



Forschung • Entwicklung • Transfer

Nürnberger Hochschulschriften
Nr. 29

Ottmann, Sebastian und König, Joachim

Was wirkt wie? – Konzeptionelle Überlegungen zur Messung
und Analyse von Wirkungen in der Sozialen Arbeit

*Der Wirkungsradar des Instituts für Praxisforschung und Evaluation der
Evangelischen Hochschule Nürnberg*

2018

Forschung, Entwicklung, Transfer – Nürnberger Hochschulschriften

Die Nürnberger Hochschulschriften zu ‚Forschung, Entwicklung, Transfer‘ sind eine lose Hochschulreihe der Evangelischen Hochschule Nürnberg zur Publikation von Vorträgen, Artikeln und Projektberichten. Auch herausragende Qualifizierungsarbeiten von Studierenden können in gekürzter Fassung veröffentlicht werden. Die Online-Reihe soll dazu dienen, das Profil der Forschung und Entwicklung an der Hochschule nach innen und außen zu kommunizieren und die breite Expertise der Mitglieder der Hochschule deutlich zu machen.

Evangelische Hochschule Nürnberg

Bärenschanzstraße 4

90429 Nürnberg

www.evhn.de

Sebastian Ottmann M.A., Prof. Dr. Joachim König

Was wirkt wie? – Konzeptionelle Überlegungen zur Messung und Analyse von Wirkungen in der Sozialen Arbeit

Nürnberg, 2018

Ottmann, S. & König, J. (2018): Was wirkt wie? – Konzeptionelle Überlegungen zur Messung und Analyse von Wirkungen in der Sozialen Arbeit. Der Wirkungsradar des Instituts für Praxisforschung und Evaluation der Evangelischen Hochschule Nürnberg. *Forschung, Entwicklung, Transfer – Nürnberger Hochschulschriften*, Nr. 29. Nürnberg: Evangelische Hochschule Nürnberg. doi: 10.17883/fet-schriften029.

Abstract

In diesem Beitrag wird der am Institut für Praxisforschung und Evaluation der Evangelischen Hochschule Nürnberg entwickelte ‚Wirkungsradar‘ vorgestellt, eine Toolbox für eine differenzierte Analyse und Messung von Wirkungen sozialer Dienstleistungen. Der Ansatz stellt ein alltagstaugliches Verfahren vor, das einen differenzierten Umgang mit dem Begriff Wirkung fordert. Bei der Umsetzung werden methodische Mindestanforderungen definiert, um eine belastbare, qualitativ hochwertige Analyse überhaupt durchführen zu können. Dazu werden sowohl Verfahren der Black-Box-Messung wie auch der White-Box-Evaluation einbezogen. Zentral im Verfahren ist ein stufenweises Vorgehen, das eine fachliche Perspektive im Rahmen der Wirkungsanalyse und das Konzept der Realistic Evaluation als Kerngedanke zugrunde legt.

Im ersten Abschnitt des Beitrages wird eine Bestimmung und Abgrenzung des Begriffs ‚Wirkung‘ vorgenommen. Danach wird Wirkungsorientierung in der Sozialen Arbeit als Thema diskutiert, bevor im vierten Abschnitt der ‚Wirkungsradar‘ detailliert dargestellt wird. Zum Abschluss werden die Grenzen von Wirkungsanalysen diskutiert, um nochmals aufzuzeigen, warum ein differenzierter Umgang mit dem Begriff nötig ist.

Inhaltsverzeichnis

1. Einleitung	5
2. Wirkung – eine Begriffsbestimmung	6
3. Wirkungsorientierung in der Sozialen Arbeit	11
4. Der Wirkungsradar als Toolbox für eine differenzierte Analyse und Messung der Wirkungen sozialer Dienstleistungen	14
4.1 Grundlage: Wirkungsworkshop.....	19
4.2 Stufe 1: Wirkmodell erstellen	19
4.3 Stufe 2: Effekte identifizieren	25
4.4 Stufe 3: Wirkung nachweisen	27
4.5 Stufe 4: Wirkmechanismen analysieren.....	31
4.6 Stufe 5: Effizienz belegen.....	33
4.7 Abschluss: Ergebnisse berichten	34
5. Grenzen der Wirkungsanalyse in der Sozialen Arbeit	35
6. Literaturverzeichnis	39

1. Einleitung

In den letzten Jahren ist die Diskussion über die Wirkung Sozialer Arbeit wieder stärker in den Blickpunkt gerückt. Gab es vor einigen Jahren schon erste Debatten hierüber in Deutschland (vgl. u. a. Eppler, Miethe & Schneider, 2011; Otto, Polutta & Ziegler, 2007), so zeigt sich, dass dieses Thema auch durch aktuelle Entwicklungen wie den Social Return on Investment (vgl. Schellberg, 2015; Schober & Then, 2015) oder die Verleihung von sogenannten „Wirkt-Siegeln“ (vgl. PHINEO gAG, 2014) wieder stärker in den Vordergrund rückt. Auch wird die Thematik aktuell erneut und intensiv innerhalb der Disziplin der Sozialen Arbeit (vgl. Borrmann & Thiessen, 2016) und unter den Wohlfahrtsverbänden (vgl. Kehl, Glänzel, Then & Mildenberger, 2016) diskutiert.

Allerdings ist dabei festzustellen, dass die Messung der Wirkungen von Interventionen und Maßnahmen schon immer Bestandteil von Evaluationen waren und es auch aktuell sind (vgl. Stockmann, 2006, S. 106). Insofern ist das Thema kein neues und es häufen sich auch Einschätzungen, dass die Frage nach dem Wirkungsnachweis innerhalb von Evaluationen in den nächsten Jahren zunehmen wird (vgl. Müller, 2017). Zukünftig wird die Frage nach der Wirksamkeit vor allem in Bereichen und Feldern der Sozialen Arbeit eine wichtige Rolle spielen, wo dies u. a. in Gesetzen, wie etwa dem neuen Bundesteilhabegesetz, explizit gefordert wird. Insofern ist es für soziale Einrichtungen, Träger und Spitzenverbände wichtig, sich mit dem Thema der Wirkung zu befassen und eine eigene Position zu diesen Fragen zu entwickeln.

Gleichzeitig zeigt sich aber auch, dass der Begriff der Wirkung in letzter Zeit nicht nur zunehmend inflationär verwendet wird, sondern auch der klare Bezug zur Fokussierung auf Kausalität, also auf eine klare Ursachen-Wirkungs-Beziehung, häufig fehlt: Die Existenz einer Wirkung wird behauptet, der empirische Nachweis ist unzureichend oder fehlt völlig. All dies hat am Institut für Praxisforschung und Evaluation der Evangelischen Hochschule Nürnberg dazu geführt, ein eigenes Konzept zur Analyse und Messung von Wirkungen zu entwickeln und damit eine ‚empirische Toolbox‘ zu beschreiben, die bei zukünftigen Projekten angewendet werden kann, in Teilen und in seinen zentralen Grundgedanken aber auch schon praktiziert wird. Zentraler Grundgedanke dieses Konzeptes ist ein differenzierter Umgang mit dem Begriff der Wirkung und seinen Implikationen. Dies erscheint nicht zuletzt auch deshalb sinnvoll, da es von einer großen Zahl von Akteuren, wie etwa auch Stiftungen, gefordert wird (vgl. Shaw & Volz, 2017).

2. Wirkung – eine Begriffsbestimmung

Laut Duden wird Wirkung als eine „durch eine verursachende Kraft bewirkte Veränderung“ (Bibliographisches Institut, o. J.) beschrieben. Durch den Bezug auf die verursachende Kraft wird hier die Wichtigkeit des kausalen Mechanismus herausgestellt. Wird von Wirkung gesprochen, muss daher klar sein, dass eine Kausalität mit dem Begriff verbunden ist. Dies hat logischerweise Konsequenzen für alle Aussagen, die im Rahmen von Wirkungsanalysen generiert werden. Im Verlauf von Forschungsprojekten ist es daher zentral, den Begriff der Wirkung von dem der Zielerreichung abzugrenzen. Diese Abgrenzung wird anhand der so genannten IOOI-Logik sichtbar: Dort werden die Inputs (Ressourcen), Outputs (Leistungen) sowie die Outcomes und Impacts (Wirkungen), als aufeinanderfolgende kausale Wirkungskette definiert (vgl. Clark, Rosenzweig, Long & Olsen, 2004). Legt man diese Logik im Rahmen einer Wirkungsanalyse zugrunde, sind die erbrachten Leistungen einer sozialen Intervention bzw. Maßnahme (z. B. die Anzahl der Beratungsstunden) noch keine Wirkung im eigentlichen Sinne. Diese werden erst im ‚nächsten Schritt‘ erzielt. Die Unterscheidung in der IOOI-Logik zwischen Outcomes und Impacts bezieht sich auf die Betrachtungsebene. So werden mit Outcomes Wirkungen auf der Ebene der Klientinnen und Klienten beschrieben, während Impacts Wirkungen auf der Ebene der Gesellschaft beschreiben (vgl. Beywl & Niestroj, 2009, S. 144)¹. Ein Beispiel für Outcomes, etwa in einem Tabak-Entwöhnungsprogramm, sind Verhaltens- oder Einstellungsänderungen bei den Klientinnen und Klienten, während ein Impact z. B. im Rückgang der Anzahl der Raucher in der Gesellschaft bestehen könnte. Diese Unterscheidung stellt natürlich auch klare Herausforderungen an eine Wirkungsanalyse, vor allem dann, wenn es darum geht, die Impacts auch wirklich zu messen. Genau aus diesem Grund findet bei den meisten Untersuchungen eine Beschränkung auf den Nachweis von Outcomes statt. Eine Auswirkung hat diese Logik aber auch schon viel früher, nämlich bei der Planung und Entwicklung von Maßnahmen, da hierbei bewusst zwischen Zielerreichung und möglichen Wirkungen, die erreicht werden sollen, unterschieden werden muss.

¹ In der Literatur zeigt sich, dass die Begriffe Outcome und Impact teilweise auch anders definiert werden: So finden sich etwa Autoren, die in einem Outcome eine kurzfristige Wirkung und im Impact eine langfristige Wirkung sehen (vgl. Zängl, 2013), aber auch Autoren die mit Impact die Wirkungen auf Klientenebene bezeichnen (vgl. Horcher & Zängl, 2013). Die Unterscheidung von Outcome als Wirkung auf Klientenebene und Impact als Wirkung auf Gesellschaftsebene erscheint jedoch insgesamt betrachtet für die Praxis der Sozialen Arbeit, gerade auch im Hinblick auf die IOOI-Logik, eindeutig zielführender.

Im Bereich der Sozialen Arbeit erscheint es aber auch sinnvoll, eine Abgrenzung von Wirkung zu anderen Resultaten sozialer Dienstleistungen wie Qualität und Nutzen vorzunehmen, da alle drei Begriffe immer wieder in der fachlichen Diskussion verwendet werden (vgl. Bleck, 2016). Während bei der Betrachtung der Wirkung die kausale Zuschreibung das zentrale Element ist, kann man Qualität als „Übereinstimmung von Leistungen mit Ansprüchen“ (Markgraf, 2018) definieren. Innerhalb des Qualitätsbegriffs wird zwischen Struktur-, Prozess- und Ergebnisqualität unterschieden und es zeigt sich, dass diese drei Ebenen in direktem Zusammenhang stehen (vgl. Bleck, 2016, S. 108 f.). In der Sozialen Arbeit ist der Qualitätsbegriff etabliert, da der Nachweis eines Qualitätsmanagements und entsprechende Zertifizierungen häufig eine Anforderung von Kostenträgern darstellt. Aufgabe des Qualitätsmanagements ist mit dem Ziel in einer Organisation verbunden, „Probleme zu erkennen und auf dieser Grundlage Strukturen und Prozesse zu optimieren“ (Wingenfeld et al., 2011, S. 8). Eine besondere Rolle in der Qualitätsdebatte, auch in Bezug auf die Abgrenzung zur Wirkung, spielt das Konstrukt der Ergebnisqualität, da sich dieses u. a. auf den Output bezieht (vgl. Merchel, 2013, S. 50). Allerdings besteht der zentrale Unterschied darin, dass bei der Erfassung von Wirkungen eben Kausalität das zentrale Kriterium darstellt und erfüllt sein muss, während bei der Ergebnisqualität lediglich betrachtet und geprüft wird, ob die geplanten Leistungen auch in entsprechender Qualität erbracht wurden. Trotzdem kann festgehalten werden, dass eine hohe Qualität – gerade auch im Bereich der Struktur- und Prozessqualität – vorhanden sein muss, damit eine Wirkung erzielt werden kann.

Neben der Frage nach Qualität wird auch häufig der Nutzen als Resultat Sozialer Arbeit definiert (vgl. Bleck, 2016, S. 113 f.). Betrachtet man dieses Konstrukt, so zeigt sich, dass hier von einer grundsätzlich anderen Sichtweise auf eine soziale Dienstleistung ausgegangen wird. Ist der Qualitätsbegriff stark von der Perspektive der Institution geprägt, wird beim Nutzen die Sichtweise der Klientinnen und Klienten in den Blick genommen. Diese Perspektive kann natürlich von der institutionellen Perspektive des Trägers abweichen und andere Aspekte und auch Resultate in den Mittelpunkt stellen. Auch wird im Rahmen der sozialpädagogischen Nutzenforschung nicht selten ein exploratives Vorgehen gewählt, das stark von qualitativen Forschungsmethoden geprägt ist. Im Hinblick auf die Debatte um den Begriff der Wirkung ist zu beobachten, dass Vertreter der sozialpädagogischen Nutzenforschung Wirkungsforschung nicht selten kritisieren (vgl. Schaarschuch & Oelerich, 2005, S. 14 ff.). An dieser Stelle sei angemerkt, dass es trotz aller methodischer Kontroversen durchaus sinnvoll und erstrebenswert erscheint, die Klientenperspektive mit in eine differenzierte Wirkungsanalyse zu integrieren. Auch wenn

Wirkungsanalysen sozialpädagogische Nutzenforschung nicht ersetzen können, ist es bei der Entwicklung von Wirkmodelle möglicherweise von entscheidendem Interesse, auch die Perspektive der Klientinnen und Klienten, sofern möglich, miteinzubeziehen, um ein vollständiges und ganzheitliches Bild möglicher Wirkungen einer Maßnahme zu erhalten.

Zusammenfassend schlägt Bleck (2016, S. 116) folgende Abgrenzung der verschiedenen Resultate Sozialer Arbeit vor:

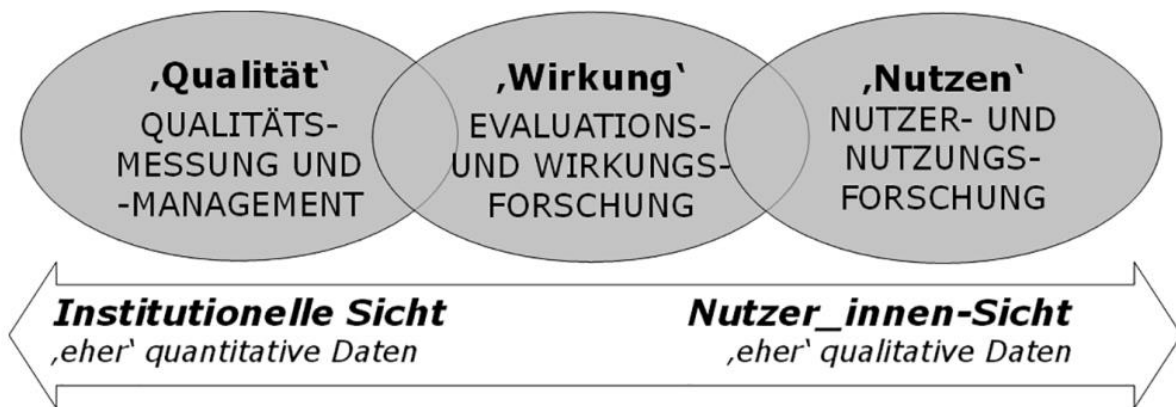


Abbildung 1: Abgrenzung von Qualität, Wirkung und Nutzen (Bleck, 2016, S. 116)

„Qualität“ zeichnet sich vor allem durch ihre institutionelle Sicht auf die Resultate Sozialer Arbeit aus und nutzt hierfür „eher“ quantitative Daten. Dem gegenüber steht der „Nutzen“, der die Perspektive der Klientinnen und Klienten in den Blick nimmt. Dazu werden in der Regel „eher“ qualitative Methoden herangezogen. Der Begriff der „Wirkung“ wird von Bleck zwischen diesen beiden Konzepten angeordnet und zeigt auch Überlappungen mit ihnen auf.

Doch was bedeuten diese Begriffsbestimmung und Einordnungen nun für das empirische Vorgehen im Rahmen einer Wirkungsanalyse? Ein differenzierter Umgang mit dem Begriff „Wirkung“ ist wichtig! So führt auch Balzer (2012, S. 138) in seinem Beitrag an, dass „hinsichtlich der Evaluation von Wirkung (...) nur Aussagen gemacht werden [dürfen], die zum Evaluationsdesign passen.“ In der Praxis zeigt sich, dass dies nicht immer der Fall ist und sehr oft allgemein von Wirkung gesprochen wird, auch wenn keine kausale Rückführung der (vermeintlichen) Wirkung auf die mögliche Maßnahme erzielt werden kann. Als Beispiel kann hier u. a. das „Wirkt-Siegel“ von Phineo genannt werden. Dieses Siegel suggeriert in der (Fach-)Öffentlichkeit, dass die untersuchte Maßnahme bzw. Organisation „wirkt“. Im Rahmen der

Analyse, die Phineo durchführt, ist dies aber gar nicht Gegenstand, sondern es geht lediglich um die Feststellung eines möglichen Wirkpotenzials. Dies bedeutet, dass nur untersucht wird, ob die Organisationsstrukturen und das Programm so aufgestellt sind, dass es wahrscheinlich ist, dass eine vorher festgelegte Wirkung erzielt wird (vgl. PHINEO gAG, 2014, S. 12).

Dieses Beispiel macht deutlich, dass ein seriöser und empirisch fundierter Umgang mit dem Begriff Wirkung wichtig ist, nicht zuletzt auch um die Akzeptanz und Glaubwürdigkeit von Wirkungsanalysen im Feld der Sozialen Arbeit zu erhöhen, zumindest aber, um sie nicht zu gefährden. Dieser Umgang ist auch im Hinblick auf die Gütekriterien empirischer Forschung absolut nötig! So ist ein zentrales Gütekriterium die Validität des Messinstrumentes, die als „Grad der Genauigkeit, mit der ein Verfahren tatsächlich das misst oder vorhersagt, was es messen und vorhersagen soll“ (Rammstedt, 2010, S. 250) definiert wird. Wie Newcomer, Hatry & Wholey (2015, S. 15) mit nachfolgender Grafik aufzeigen, kann eine aussagekräftige Evaluationsstudie – und damit auch eine Wirkungsanalyse – nur erfolgen, wenn die Validität der Messinstrumente gegeben ist.

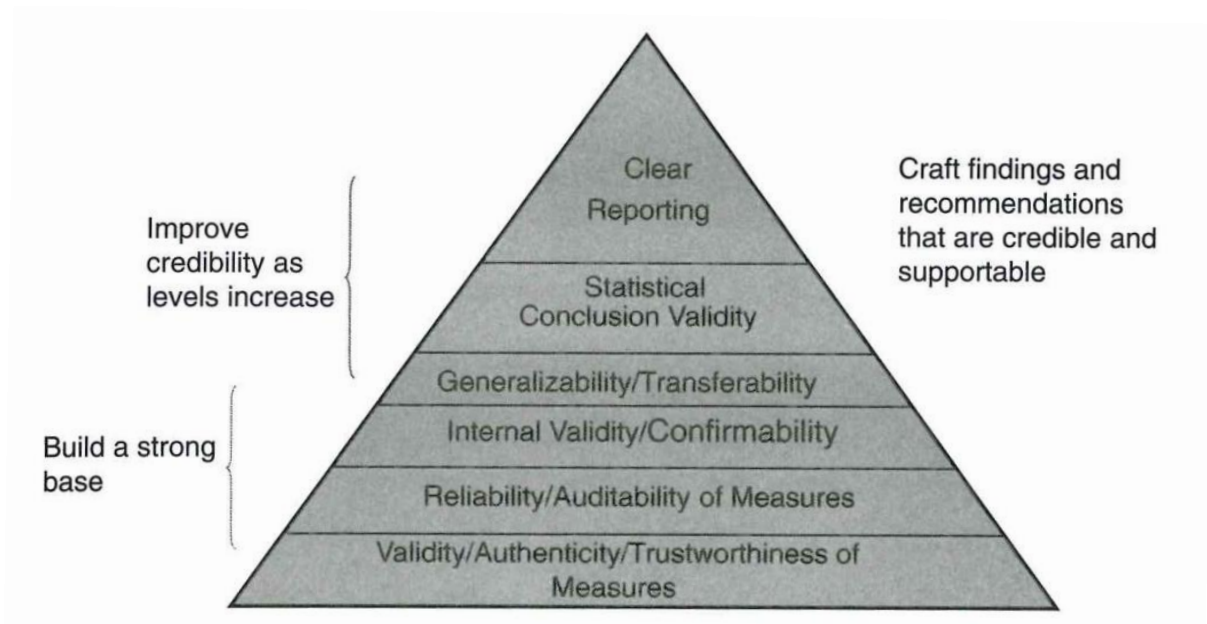


Abbildung 2: Strength of Evaluation (Newcomer et al., 2015, S. 15)

Weiterhin ist die Sicherstellung von Validität und Reliabilität der gewonnen Informationen auch in den Standards für Evaluationen (vgl. DeGEval – Gesellschaft für Evaluation e. V., 2017, S. 46) festgeschrieben, die bei der Anwendung von Projekten am Institut für Praxisforschung und Evaluation berücksichtigt und eingehalten werden. Genau aus diesem Grund wird bei dem hier

vorgeschlagenen Instrumentarium, dem ‚Wirkungsradar‘, ein differenzierter Umgang mit dem Begriff zu Grunde gelegt. Um zum Ausdruck zu bringen, dass eine Wirkung eben häufig nicht ohne Einschränkungen kausal auf eine Maßnahme zurückgeführt werden kann, wird im Weiteren mit dem Begriff der „Wirkungsplausibilisierung“ gearbeitet, den Balzer (2012, S. 136) eingeführt hat. Balzer versteht darunter alle empirischen Designs, die keinen kausalen Wirknachweis erbringen (können). Dies können u. a. Veränderungsmessungen oder theoriebasierte Ansätze sein. Zentral bei diesem Vorgehen ist, dass bei der Dateninterpretation auch Vermutungen angestellt werden sollen, ob die gefundenen Veränderungen auf die Maßnahme zurückzuführen sind, bzw. welche weiteren Einflussfaktoren es für diese gibt oder geben könnte. Der Begriff der Wirkungsplausibilisierung sollte daher auch bei der Kommunikation von Ergebnissen genutzt werden, um eine solide und um sachliche Klarheit bemühte Differenzierung in der Wirkungsdebatte zu erreichen.

Eine weitere, notwendige Differenzierung, die einen wichtigen Einfluss auf empirische Forschungsdesigns im Zusammenhang mit Wirkungsanalysen hat, ist die Unterscheidung in Brutto- und Nettowirkung (vgl. Caspari, 2009 und Abbildung 3). In der Bruttowirkung sind alle Veränderungen zwischen zwei Zeitpunkten berücksichtigt, beispielsweise zwischen dem Beginn und dem Ende einer Maßnahme. Aus diesem Grund ist die Bruttowirkung auch keine Wirkung im eigentlich Wortsinn, da nicht festgestellt werden kann, welchen kausalen Einfluss die untersuchte Maßnahme bzw. soziale Intervention auf die Wirkung hat. Daher sollte im Rahmen von Wirkungsanalysen die Messung der so genannten Nettowirkung im Mittelpunkt stehen. Hier wird nur der Anteil der Veränderung zwischen den zwei Zeitpunkten in den Blick genommen, die auf die Maßnahme tatsächlich zurückzuführen ist. Auch die Betrachtung der Neben- und Folgewirkungen spielt dabei eine wichtige Rolle. Für ein mögliches Forschungsdesign bedeutet dies, dass neben einem Längsschnitt (mind. zwei Zeitpunkte) und der Maßnahmengruppe auch eine weitere Vergleichsgruppe in die Analyse miteinbezogen werden muss, um einen Wirknachweis empirisch valide erbringen zu können.

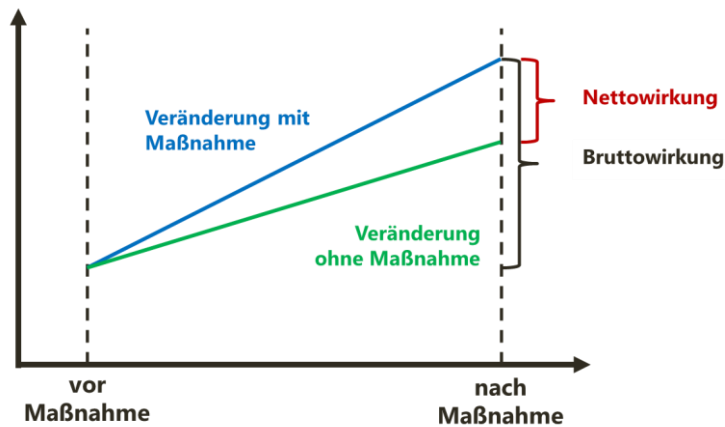


Abbildung 3: Unterschied zwischen Brutto- und Nettowirkung (in Anlehnung an Caspari, 2009, S. 191)

Abschließend kann daher festgestellt werden, dass der differenzierte Umgang mit dem Begriff „Wirkung“ zentral und unerlässlich für jede Wirkungsanalyse ist. Bei der Darstellung der Ergebnisse sollte demnach nur dann von Wirkung gesprochen werden, wenn mit einer hohen Wahrscheinlichkeit die gefundenen Veränderungen auch kausal auf die Intervention, also das Angebot der Sozialen Arbeit zurückzuführen sind. Ist dies nicht der Fall, sollte eine Abstufung erfolgen. Dazu kann der Begriff der ‚Wirkungsplausibilisierung‘ genutzt bzw. von Veränderungen oder Effekten auf der Ebene der Klientinnen und Klienten gesprochen werden. Wichtig erscheinen dann auch Hinweise, dass die dargestellten Befunde auch von anderen Außeneinflüssen bedingt sein könnten und die Intervention somit nicht als alleiniger Auslöser gelten kann.

Nachdem der Begriff der Wirkung nun bestimmt und daran anschließend ausgeführt wurde, welche Konsequenzen dies für die entsprechenden Forschungsdesigns hat, soll im nächsten Abschnitt kurz auf die grundsätzliche Debatte um Wirkungsorientierung in der Sozialen Arbeit eingegangen werden.

3. Wirkungsorientierung in der Sozialen Arbeit

Die Debatte über die Wirkungen von Angeboten der Sozialen Arbeit ist eingebettet in die Diskussion über eine mögliche Wirkungsorientierung der Sozialen Arbeit. Dies ist keine neue Diskussion, denn der Begriff der Wirkungsorientierung taucht schon Anfang der 1990er Jahre durch die Einführung des New Public Managements, bzw. in Deutschland durch den Begriff der ‚Neuen Steuerung‘ auf (vgl. Horcher, 2013; Polutta, 2013). Von Wirkungsorientierung ist dort

immer „dann die Rede, wenn eine Hinwendung zu Wirkungen sozialer Dienste gefordert wird“ (Polutta, 2013, S. 1108). Diese Definition sollte allerdings noch erweitert werden, da in der aktuellen Debatte der Begriff der Wirkungsorientierung auch dann Anwendung findet, wenn eine soziale Organisation ihre Angebote und Leistungen auf die Erzielung einer gewünschten Wirkung aktiv ausrichtet. Dieser Aspekt kam in den Anfangsjahren der Debatte zu kurz. So wurde u. a. Wirkungsorientierung lediglich als ein Instrument der Finanzierung sozialer Dienste gesehen. Die Perspektiven der evidenzbasierten Praxis, also die Integration von Forschungsergebnisse in das fachliche Handeln, wurden dagegen nur sehr wenig betrachtet (vgl. Polutta, 2013). Aus diesem Grund spricht auch Otto (2007, S. 22 ff.) in dem Zusammenhang von einem „halbierten Wirkungsdiskurs“. In den letzten Jahren zeigt sich aber, dass unter dem Schlagwort ‚Wirkungsorientierung‘ auch vermehrt die Integration von möglichen Wirkungen und Forschungsergebnissen in die eigene Arbeit, d. h. in das Handeln der pädagogischen Fachkräfte zu beobachten ist, auch wenn es an dem Konzept immer wieder auch Kritik gibt, z. B. dass verpflichtende Maßnahmen der Wirkungsorientierung keinen direkten Nutzen für die Klientinnen und Klienten haben (vgl. Kratz, 2017).

Das Konzept der Wirkungsorientierung wird sowohl auf der Ebene der Kostenträger bzw. der Politik als auch auf der Ebene der sozialen Einrichtungen umgesetzt: Bei den Kostenträgern und in der Politik zeigt sich dies vor allem durch die zuvor schon benannte Einführung der sog. Neuen Steuerung. Aber auch aktuell werden Projekte durchgeführt, um kommunale Steuerung wirkungsorientiert zu gestalten (vgl. PHINEO gAG, 2018). In der aktuellen Gesetzgebung zeigt sich ebenfalls, dass ein Nachweis der Wirksamkeit von Leistungen von der Politik immer stärker eingefordert wird, z. B. im neuen Bundesteilhabegesetz (BTHG; vgl. § 121 Abs. 2 SGB IX-neu und § 125 Abs. 1 SGB IX-neu). Diese Bestrebungen innerhalb der Politik und bei den Kostenträgern haben natürlich auch Auswirkungen auf die Erbringer von sozialen Dienstleistungen. Wird von Seiten der Politik und der Kostenträger eine wirkungsorientierte Steuerung bzw. ein Wirkungsnachweis eingefordert, so befinden sich Einrichtungen und Träger der Sozialen Arbeit unter Zugzwang: Eine frühzeitige Auseinandersetzung mit dem Thema erscheint sinnvoll, nicht nur um bei möglichen Diskussionen eigene Standpunkte einbringen zu können.

Betrachtet man die Felder der Sozialen Arbeit, so zeigt sich, dass das Thema Wirkungsorientierung in den verschiedenen Bereichen unterschiedlich stark ausgeprägt ist. Auf der einen Seite sind Arbeitsfelder zu beobachten, die sich schon sehr intensiv mit dem Thema beschäftigt haben, wie

etwa die Kinder- und Jugendhilfe (vgl. u.a. Albus et al., 2010; Renner, 2012), möglicherweise auch deshalb, weil dies in den zugrundeliegenden Gesetzen im Sinne einer Wirkungsforschung gefordert wird, wie etwa in der Arbeitsmarktpolitik (vgl. Polutta, 2013, S. 1109). Auf der anderen Seite sehen wir Arbeitsfelder, in denen das Thema gerade erst an Aktualität gewinnt, beispielsweise im Bereich der Behindertenhilfe.

Im Hinblick auf die Durchführung von Wirkungsanalysen bleibt festzuhalten, dass diese immer in einem Prozess der Wirkungsorientierung integriert sein sollten, also nicht ‚frei schwebend‘ durchgeführt werden können. Nachfolgende Darstellung stellt einen in dieser Hinsicht sinnvoll erscheinenden Implementierungskreislauf der Wirkungsorientierung in sozialen Einrichtungen dar:

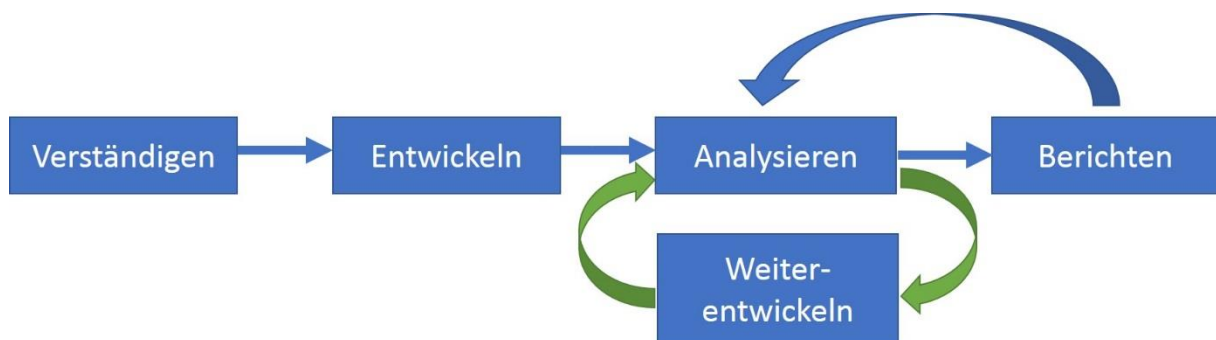


Abbildung 4: Implementierungskreislauf im Rahmen der Wirkungsorientierung

In einem ersten Schritt muss eine Verständigung innerhalb der Einrichtung erzielt werden, was unter dem Begriff der Wirkung zu verstehen ist (siehe auch Punkt 2 in diesem Beitrag). Die Umsetzung der Wirkungsorientierung in der Einrichtung findet dann bereits vor einer möglichen Wirkungsanalyse statt, indem die Angebote und Maßnahmen im Hinblick auf die gewünschten, also zu erzielenden Wirkungen entwickelt und ausgestaltet werden. Die Wirkungsanalyse dient dazu, zu überprüfen, ob die zuvor festgelegten Wirkungen und Veränderungen erzielt werden und bietet neben der Legitimation durch Berichte vor allem auch die Möglichkeit, Ansatzpunkte für die Weiterentwicklung des untersuchten Angebotes aufzuzeigen und konzeptionell (weiter) zu entwickeln.

Daher ist das Konzept der Wirkungsorientierung nicht nur ein Instrument für die Finanzierung von Einrichtungen – wobei es durchaus legitim erscheint, wenn Politik und Kostenträger vor allem solche Angebote finanzieren möchten, die eine zuvor definierte Wirkung erzielen –, sondern auch

ein wichtiges Instrument zur Organisations-, Konzeptions- und Angebotsentwicklung in sozialen Einrichtungen. Die wirkungsorientierte Ausrichtung von Einrichtungen der Sozialen Arbeit sollte vor allem als Chance verstanden werden, Angebote in diesem Feld weiterzuentwickeln und eine empirische Wissensbasis für die Durchführung von Angeboten und Maßnahmen der Sozialen Arbeit – letztlich immer zum Nutzen der Klientinnen und Klienten – aufzubauen.

4. Der Wirkungsradar als Toolbox für eine differenzierte Analyse und Messung der Wirkungen sozialer Dienstleistungen

Mit dem ‚Wirkungsradar‘, der als empirische Toolbox am Institut für Praxisforschung und Evaluation der Evangelischen Hochschule Nürnberg entwickelt wurde, soll ein flexibles Verfahren zur Verfügung gestellt werden, mit dem eine differenzierte und alltagstaugliche Wirkungsanalyse möglich ist. Gleichzeitig werden aber auch methodische Mindestanforderungen definiert, um eine qualitativ hochwertige Analyse überhaupt durchführen zu können. Hierbei werden sowohl Verfahren der Black-Box-Messung wie auch der White-Box-Evaluation vereint. D. h. im Rahmen der Analyse mit Hilfe des ‚Wirkungsradars‘ wird nicht nur gefragt, *ob* ein Angebot wirkt (Black-Box), sondern auch *wie* diese Wirkungen zustande kommen (White-Box). Die grundlegenden und methodisch zentralen Annahmen des Instruments sind:

- Ein stufenweises Vorgehen.
- Das Zugrundlegen einer fachlichen Perspektive im Rahmen der Wirkungsanalyse.
- Das Konzept der Realistic Evaluation als Grundgedanke, mit dem Ziel der Entwicklung von Theorien mittlerer Reichweite.

Nachfolgend werden diese zentralen Annahmen kurz dargestellt, bevor dann der ‚Wirkungsradar‘ selbst in seinem Aufbau näher dargestellt wird.

Das *stufenweise Vorgehen bei der Analyse von Wirkungen* sozialer Dienstleistungen soll sicherstellen, dass zum einen ein auf den Untersuchungsgegenstand abgestimmtes Vorgehen möglich ist, zum anderen auch ein differenzierter Umgang mit dem Begriff Wirkung erfolgen und dargestellt werden kann, ob ein empirisch valider Wirknachweis vorliegt oder ob nur Effekte zur Wirkungsplausibilisierung gemessen wurden. Zudem orientiert sich das stufenweise Vorgehen an sozialwissenschaftlich evidenten Forderungen (vgl. Veerman & van Yperen, 2007) und wird u. a.

auch in der ‚Grünen Liste Prävention‘ in diesem Sinne umgesetzt (vgl. Landespräventionsrat Niedersachsen, 2011). So kann durch die Anwendung des ‚Wirkungsraders‘ eine Anschlussfähigkeit hergestellt werden.

Ein weiterer wichtiger Aspekt des stufenweisen Vorgehens für Träger und Einrichtungen ist, dass eine Befassung mit dem Thema Wirkungsorientierung und Wirkungsanalyse auf einem zunächst einfach überschaubaren Niveau gestartet werden kann und danach – darauf aufbauend – die Möglichkeit für tiefergehende Analysen besteht. Aus diesem Grund wird die Toolbox auch als alltagstauglich in der Umsetzung angesehen. Das stufenweise Vorgehen hat zudem den Vorteil, dass ein umfassender Wirkungsnachweis erst dann erbracht werden muss, wenn auch tatsächlich erste Anhaltspunkte für Veränderungen bei den Klientinnen und Klienten festgestellt wurden (vgl. Balzer, 2012, S. 135). Die einzelnen Stufen des Wirkungsraders könnten nachfolgender Abbildung entnommen werden:

Der IPE Wirkungsradar

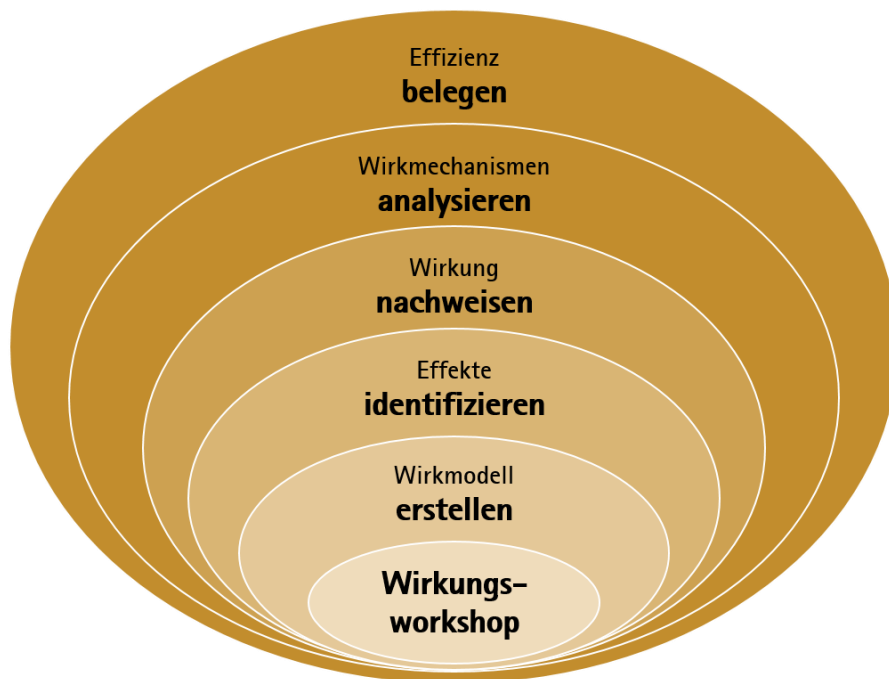


Abbildung 5: IPE Wirkungsradar

Mit dem Wirkungsradar soll – zweite Grundannahme – eine *fachliche Perspektive* auf die Wirkungsanalyse erfolgen. Im Vordergrund steht dabei das Ziel, neben dem Wirkungsnachweis auch Anhaltspunkte für die Weiterentwicklung der untersuchten Maßnahme sowie theoriebildendes Wissen für die eigene Praxis zu erhalten. Erst in einem letzten Schritt wird im Rahmen der Untersuchung die Effizienz der untersuchten Maßnahme bzw. des Programms belegt. Vor dem Hintergrund dieser fachlichen Perspektive der Analyse ist es erfahrungsgemäß sinnvoll und Gewinn bringend, dass in der Einrichtung, die eine Wirkungsanalyse durchführt bzw. beauftragt, dazu eine (kleine) Begleitgruppe einrichtet. Diese kann u. a. bei der Entwicklung eines Wirkmodells sowie bei der Bewertung der gefundenen Ergebnisse entscheidende Beiträge aus der Perspektive der jeweiligen Praxis leisten.

Als grundlegendes Evaluationsverständnis findet, so die dritte Grundannahme, die *Realistic Evaluation* (vgl. Pawson & Tilley, 1997) Anwendung. Zentral ist in diesem Ansatz, dass das zugrunde gelegte Kausalitätsverständnis sowohl um den Kontext als auch um die Mechanismen erweitert wird und beides dann auch im Rahmen der Evaluation in den Blick genommen wird. Hierbei werden sogenannte KMO-Konfigurationen konstruiert: Kontext + Mechanismus = Outcome.

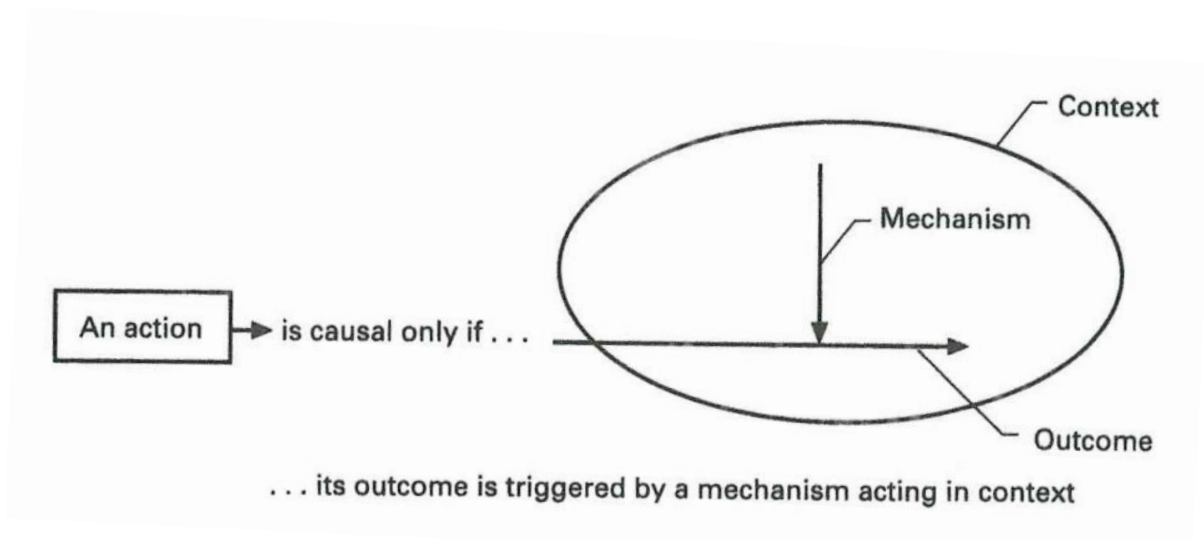


Abbildung 6: Kontext, Mechanismen und Outcomes in der Realistic Evaluation (Pawson & Tilley, 1997, S. 58)

Im Zuge dieses Ansatzes wird also folgende Logik einer realistischen Erklärung zugrunde gelegt: „Ein Programm bzw. eine Intervention erzeugt nur dann Wirkungen, wenn diese in einem konkreten Kontext Wirkungszusammenhänge aktivieren, die sich beobachten lassen und Veränderungen erzeugen“ (Haunberger & Baumgartner, 2017, S. 126). Die Berücksichtigung des Kontexts und der Mechanismen erscheint gerade im Bereich der Sozialen Arbeit deshalb besonders sinnvoll, weil bei Maßnahmen und Programmen in diesem Bereich erfahrungsgemäß eine Vielzahl von Außeneinflüssen vorhanden sind, die eine Wirkung auf die Klientinnen und Klienten haben können. Die umfassende Analyse des Kontextes muss zudem auch deshalb zwingend erfolgen, weil sie Bestandteil der Standards für Evaluation (vgl. DeGEval – Gesellschaft für Evaluation e. V., 2017, S. 44 f.) ist, die bei der Anwendung des ‚Wirkungsraders‘ Berücksichtigung finden.

Durch dieses erweiterte Verständnis ist der ‚Wirkungsradar‘ (wie auch die Realistic Evaluation) in der Lage, sogenannte Theorien mittlerer Reichweite (vgl. Merton, 1968; Pawson, 2010) als Ergebnis hervorzubringen. Theorien mittlerer Reichweite haben „keinen Anspruch auf Generalisierung, gehen aber über Aussagen zu einem Einzelfall hinaus“ (Haunberger & Baumgartner, 2017, S. 125). In der Konsequenz bedeutet dies, dass gefundene Wirkungen bei der Analyse einer Maßnahme nicht unbedingt bedeuten, dass diese bei der Implementierung in einem anderen Kontext auch erzielt werden können. Dieser Aspekt muss natürlich auch bei der Kommunikation der Ergebnisse von Wirkungsanalysen berücksichtigt werden.

Für ein mögliches Forschungsdesign gibt es im Rahmen der Realistic Evaluation keine Vorgaben. Eine systematischer Literaturreview von Haunberger & Baumgartner (2017) zur Anwendung der Realistic Evaluation in der Sozialen Arbeit kommt aber zu dem Ergebnis, dass bei den in diesem Zusammenhang analysierten Untersuchungen überwiegend mit qualitativen Methoden sowie mit Mixed-Methods-Design gearbeitet wird (vgl. S. 136). Im Rahmen des Wirkungsraders sollten daher im Evaluationsrahmen der Realistic Evaluation ebenfalls quantitative wie auch qualitative Methoden eingesetzt werden. So wird der eigentliche kausale Wirkungsnachweis (Stufe 3) quantitativ, verbunden mit einem Vergleichsgruppendesign und einer Vorher-Nachher-Messung, erbracht. Liegen diese Voraussetzungen jedoch nicht vor, kann auch nicht von einem Wirkungsnachweis im eigentlichen Sinne ausgegangen und gesprochen werden. Bei der Entwicklung von Wirkmodellen (Stufe 1) und der Analyse von Wirkmechanismen (Stufe 4) kommen in der Regel immer auch qualitative Methoden zum Einsatz.

Bevor die einzelnen Stufen des ‚Wirkungsraders‘ näher beschrieben werden, soll noch auf die vielfältigen Arten, Formen und Reichweiten von Wirkungen sowie auf mögliche Einsatzgebiete des Instrumentariums in diesem Zusammenhang eingegangen werden. Rauscher, Mildnerberger & Krlev (2015, S. 48) systematisieren die verschiedenen Arten in der nachfolgenden Abbildung auf eine sehr übersichtliche und informative Art und Weise:

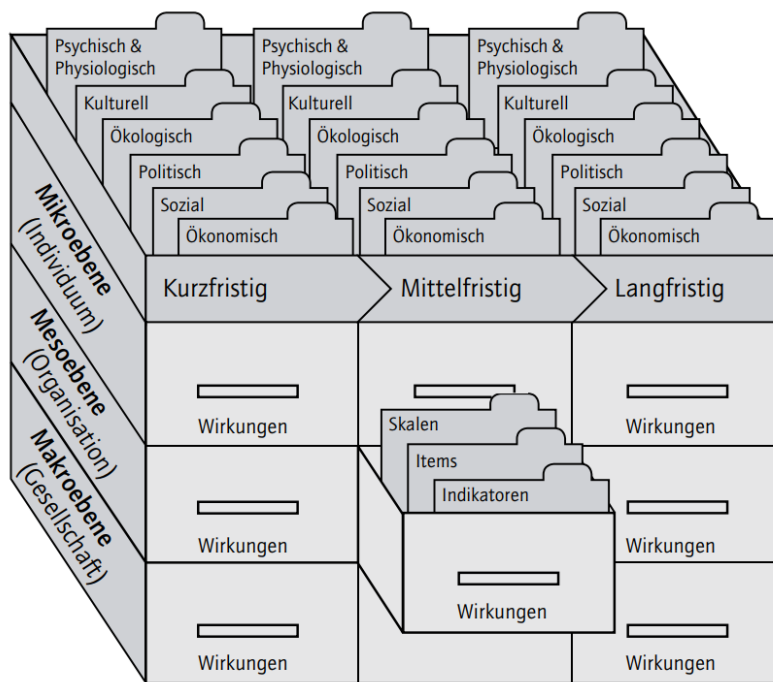


Abbildung 7: Arten von Wirkungen (Rauscher et al., 2015, S. 48)

So können Wirkungen sowohl auf der Mikro- und Meso-, wie auch auf der Makroebene erzielt werden. Der ‚Wirkungsradar‘ nimmt zunächst die Wirkungen auf der Mikroebene bei den Klientinnen und Klienten, die im Rahmen von Angeboten der Sozialen Arbeit (einschließlich der Pflege) betreut und begleitet werden, in den Blick. Hierbei werden auch Wirkungen auf Multiplikatorenebene mitbetrachtet (siehe Abschnitt 4.2). Wollte man im Rahmen der Untersuchung darauf aufbauend auch die Wirkungen auf der Meso- und Makroebene genauer analysieren, so kann der ‚Wirkungsradar‘ jederzeit um weitere Teilstudien ergänzt werden. Allerdings ist die Erfassung von gesellschaftlichen Wirkungen auf Ebene von einzelnen Angeboten und Leistungen schwierig und sollte daher auf „höherer Ebene“ oder mit aggregierten Daten erfasst werden.

Im Rahmen von Wirkungsanalysen zeigt sich, dass diese vor allem bei klar abgegrenzten Maßnahmen und Programmen ohne Probleme angewendet werden können. Soll die Wirkung von stationären Angeboten oder von Gesamtplanverfahren analysiert werden, wird der Forschungsgegenstand jedoch in aller Regel komplexer und es muss gut überlegt werden, welches Vorgehen hier gewählt wird. Beim Einsatz des ‚Wirkungsraders‘ in einer solchen Untersuchung werden in einem ersten Schritt die zentralen Angebote der stationären Einrichtung bzw. die Leistungen, die in einem Gesamtplanverfahren enthalten sind, identifiziert. Danach kann der ‚Wirkungsradar‘ auf diese einzelnen Angebote angewendet werden, um mögliche Effekte und Wirkungen zu identifizieren. Nach diesen Analysen wird abschließend eine übergreifende Untersuchung durchgeführt, in der die Koordination und die Stimmigkeit der verschiedenen Angebote in den Blick genommen werden². Durch dieses Vorgehen können dann auch valide Aussagen über Effekte und Wirkungen von stationären Einrichtungen bzw. Gesamtplanverfahren erzielt werden.

4.1 Grundlage: Wirkungsworkshop

Wird in einer Einrichtung zum ersten Mal eine Wirkungsanalyse mit dem ‚Wirkungsradar‘ durchgeführt, erfolgt als Grundlage zunächst ein Wirkungsworkshop. Ziel des Workshops ist die Vermittlung des grundsätzlichen Verständnisses von Wirkung und Wirkungsorientierung sowie die Erarbeitung einer eigenständigen Position der verantwortlichen Fachkräfte in der Einrichtung zu diesem Thema. Aus diesem Grund sollten neben Führungskräften auch Mitarbeitende, die im täglichen Kontakt mit Klientinnen und Klienten stehen, daran teilnehmen. Das erarbeitete Verständnis fließt zum einen in die weiterführenden Analysen ein, kann aber auch im Rahmen der fachpolitischen Diskussion genutzt werden und ermöglicht es der Einrichtung bereits an dieser Stelle, sprachfähig zu diesem Thema zu werden.

4.2 Stufe 1: Wirkmodell erstellen

Ziel dieser Stufe ist die Entwicklung eines Wirkmodells, das auch als Programmtheorie bezeichnet werden kann und die Grundlagen für alle weiteren Analysen legt. Ein Wirkmodell kann man als „eine logische, grafische aufbereitete Darstellung dessen, wie ein Programm [oder eine

² Ein ähnliches Vorgehen wird auch bei komplexen Programmen in der Entwicklungszusammenarbeit diskutiert (vgl. Bamberger, Vaessen & Raimondo, 2016).

Maßnahme] unter gewissen Rahmenbedingungen theoretisch und/oder empirisch funktioniert" (Rauscher et al., 2015, S. 43) verstehen. Im Rahmen des ‚Wirkungsraders‘ wird als Methode zur Entwicklung eines solchen Modells die Theory of Change (vgl. Fulbright-Anderson, Kubisch & Connell, 1998) angewendet. Dazu wird ein logisches Modell (Logic Model) des Programms bzw. der Maßnahme als ein Ergebnis des Prozesses gesehen (vgl. Harries, Hodgson & Noble, 2014). Einzelne Autoren sehen in der Theory of Change und dem Logic Model zwei unterschiedliche Ansätze (vgl. u. a. Grinnell, Gabor & Unrau, 2016; Knowlton & Phillips, 2013), diese Unterscheidung wird jedoch in der praktischen Umsetzung als nicht praktikabel angesehen. Festzustellen ist, dass die Theory of Change, im Vergleich zu einem Wirkmodell, die umfassendere Methode ist und bereits zur Programm- bzw. Maßnahmenplanung eingesetzt werden kann (vgl. Rauscher et al., 2015, S. 51).

Zentrales Element der Theory of Change ist das sogenannte „backward mapping“. Dazu ist das Abarbeiten der folgenden Schritten notwendig (vgl. The Center for Theory of Change, o. J.):

- In einem ersten Schritt wird die ‚Hauptwirkung‘ der Maßnahme festgelegt.
- Danach wird durch eine Rückwärtsinduktion herausgearbeitet, welche Zwischenresultate nötig sind, um diese ‚Hauptwirkung‘ zu erzielen. Hierbei wird der Blick zunächst auf die Outcomes, also die Wirkungen, gerichtet und noch nicht auf die Aktivitäten selbst. Es werden also ‚Teilwirkungen‘ definiert, die zur Erreichung der ‚Hauptwirkung‘ vorhanden sein müssen. Daher empfiehlt es sich dazu, eine stufenweise Definition der Outcomes im Sinne einer sogenannten Resultatetreppe (vgl. Beywl & Niestroj, 2009, S. 145 und Abbildung 8) vorzunehmen: Die Unterscheidung von Outcomes im Sinne einer Aneignung von Wissen und Fähigkeiten einerseits und der Veränderung von Handlungen und Verhalten im nächsten Schritte können wichtige Anhaltspunkte bei der Ausarbeitung eines detaillierten Wirkmodells darstellen.

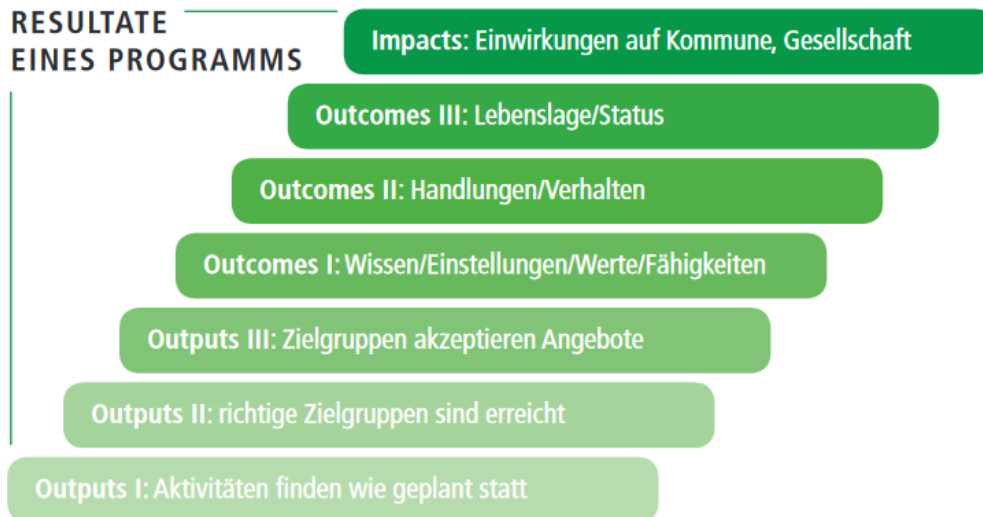


Abbildung 8: Resultate eines Programms (Beywl & Niestroj, 2009, S. 145)

- Im dritten Schritt werden Annahmen über den Kontext, in den die Maßnahme eingebettet ist, erarbeitet. Unter Kontext werden u. a. gesellschaftliche, finanzielle und/oder rechtliche Rahmenbedingungen verstanden.
- Danach werden die Aktivitäten identifiziert, die durchgeführt werden müssen, damit die zuvor definierten Teil- und Hauptwirkungen erzielt werden können. In Bezug auf die zuvor definierten Outcomes kann es solche Wirkungen geben, die von „alleine“ entstehen und solche, bei denen eine Intervention im Sinne der Maßnahme nötig ist.
- In einem letzten Schritt werden dann auch schon bei der Erstellung der „Theory of Change“ Indikatoren festgehalten, die später zur Überprüfung der Umsetzung der Maßnahmen dienen können.

Am Ende des Prozesses wird das erstellte logische Modell visualisiert und verschriftlicht. Da im Rahmen der Theory of Change der Blick vor allem auf die Wirkungen, den Kontext und die Aktivitäten gerichtet ist, sollte im Rahmen des Wirkungsradars bei der Darstellung des Wirkmodells eine Ergänzung durch den Programmbaum (vgl. Bartsch, Beywl & Niestroj, 2016 und Abbildung 9) erfolgen.

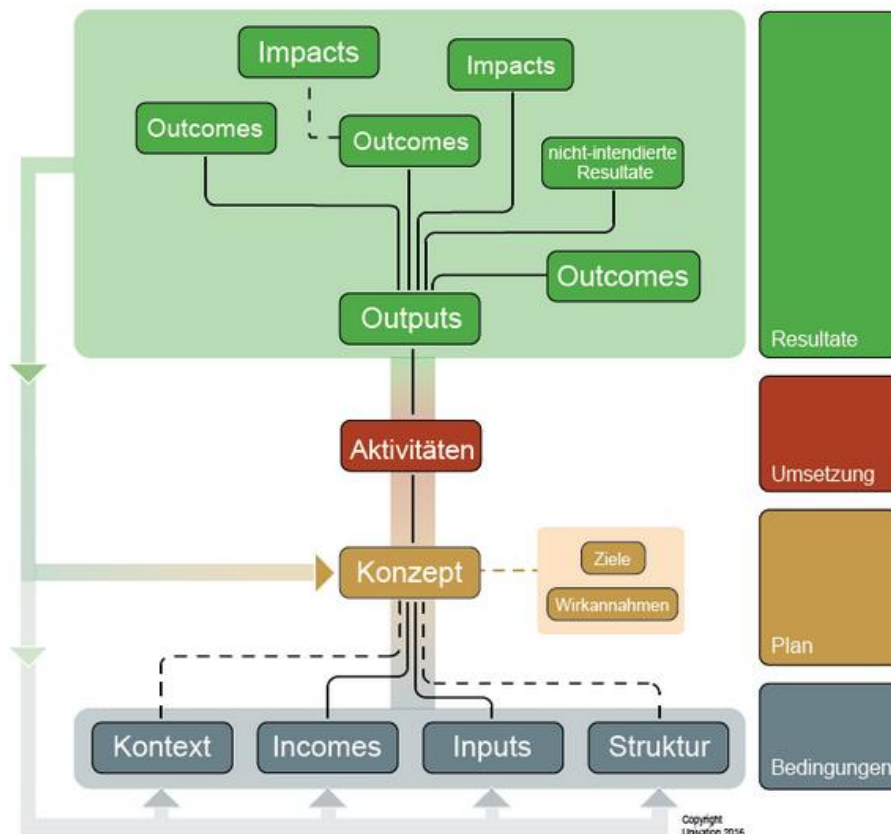


Abbildung 9: Programmbaum (Bartsch et al., 2016)

Durch diese Ergänzung werden auch die Incomes, die Inputs und die Struktur genauer beschrieben. Die Beschreibung dieser Bedingungen wird im Rahmen des ‚Wirkungsraders‘ als essentiell für eine erfolgreiche Umsetzung und Analyse von Programmen und Maßnahmen angesehen. Die Incomes beschreiben Ressourcen, die die Klientinnen und Klienten mit einbringen. Diese können z. B. das Vorwissen oder die Motivation bei Jugendlichen in Maßnahmen der Arbeitsweltbezogenen Jugendsozialarbeit sein. Unter Inputs werden alle Ressourcen beschrieben, die die Einrichtung und Träger einsetzt. Ebenfalls auf die Einrichtung bezieht sich die Struktur. Hierunter werden Bedingungen verstanden, die zur erfolgreichen Umsetzung der Maßnahme vorliegen müssen. Nach Bartsch et al. (2016) werden hierunter auch Bedingungen bei Kooperationspartnern, die für die Durchführung der Maßnahme nötig sind, beschrieben.

Zusätzlich zu den genannten Elementen erscheint es auch sinnvoll, bei der Entwicklung von Wirkmodellen auch gleich mögliche Außeneinflüsse auf die gewünschte Wirkung mit zu beschreiben. Dadurch ist sichergestellt, dass diese bei der nachfolgenden Analyse berücksichtigt werden können. Weiterhin muss geprüft werden, ob im Rahmen des Wirkmodells Aussagen über die Persistenz und den Transfer des Erlernten erfolgen sollen. Unter Persistenz versteht man die

Nachhaltigkeit einer Maßnahme, also ob das Erlernte über die Maßnahme hinaus wirkt. Unter Transfer wird verstanden, dass die Teilnehmenden das erlernte Wissen bzw. das veränderte Verhalten auch auf andere Situationen übertragen können (vgl. Gollwitzer & Jäger, 2014, S. 113 ff.).

Durch das Zusammenspiel von ‚Theory of Change‘ und dem Ansatz des Programmbaums können nach allen Erfahrungen aussagekräftige Programmtheorien und Wirkmodelle erstellt werden, die auch den Kontext, in denen die Programme und Maßnahmen eingebettet sind, gut und umfassend beschreiben. Hierdurch ist es zudem möglich, in diesem Prozess die benötigten KMO-Konfigurationen im Rahmen der Realistic Evaluation zu erstellen.

Um mit dem Ansatz der Theory of Change arbeiten zu können, muss innerhalb der Einrichtung bereits eine Vorstellung über eine Hauptwirkung einer Maßnahme oder eines Programms vorhanden sein. Aus Sicht und Erfahrung der Autoren ist dies auch in aller Regel der Fall. Falls nicht, wird in einer Vorstufe mit der Methode des ‚Problembaums‘ gearbeitet, bevor der Ansatz der Theory of Change umgesetzt wird. Der ‚Problembaum‘ ist ein Instrument aus dem Ansatz des ‚Logical Framework Approach‘. Bei der Erstellung eines ‚Problembaums‘ werden partizipativ die negativen Aspekte der aktuellen Situation gesammelt und in eine Ursachen-Wirkung-Beziehung gebracht. Im nächsten Schritt wird der erstellte ‚Problembaum‘ in einen ‚Zielbaum‘ (objectiv tree) weiterentwickelt, in dem die gewünschte zukünftige Situation beschrieben wird. Durch diesen ‚Zielbaum‘ wird u. a. auch eine Hauptwirkung definiert, die dann in einen möglichen Theory of Change-Prozess einfließen kann. Weiterhin zeigt der ‚Zielbaum‘ auch verschiedene mögliche Wege zur Erreichung des Ziels auf und man kann sich für den Weg entscheiden, der am besten umzusetzen ist (vgl. SECO, 2017, S. 8 ff.).

Des Weiteren muss bei der Erstellung des Wirkmodells geprüft werden, ob eine Differenzierung nach verschiedenen Stakeholdern nötig ist (vgl. Rauscher et al., 2015, S. 43 f.). Dazu muss klar definiert werden, an wen sich die untersuchten Maßnahmen bzw. das Programm richten und ob noch weitere Stakeholder involviert sind. So kann z. B. eine Unterscheidung zwischen Multiplikatorenzielgruppe und Letztzielgruppe sinnvoll erscheinen, oder ein Programm richtet sich sowohl an Kinder wie auch an Eltern und soll innerhalb dieser Zielgruppen unterschiedliche Wirkungen entfalten (vgl. Schmidt, 2016, S. 171 ff.).

In der praktischen Erarbeitung eines Wirkmodells kann im Rahmen des ‚Wirkungsradars‘ mit verschiedenen Methoden gearbeitet werden:

- In einem ersten Schritt werden vorhandene Konzeptionen und Dokumente zum Untersuchungsgegenstand analysiert. Weiterhin wird überprüft, inwieweit es theoretische Zugänge gibt, die inhaltlich zur untersuchten Maßnahme bzw. zum Programm passen. Ein möglicher Ansatz, der hier zur Anwendung kommen könnte, ist der Capability Approach (Befähigungsansatz) der auf Amanda Sen zurück geht und der u. a. als Grundlage definiert, dass Menschen über Befähigungen verfügen müssen um das eigene Leben erfolgreich zu gestalten (vgl. u.a. Sedmak, Babic, Bauer & Posch, 2011).
- Mithilfe von Fokusgruppen und Leitfadeninterviews können Aspekte für das spätere Wirkmodell erarbeitet werden. Diese sollten mit allen identifizierten Stakeholdern durchgeführt werden. Wichtig ist hier auch die Sicht der Klientinnen und Klienten, die nach Möglichkeit mit erfragt und berücksichtigt wird.
- Weiterhin können in diesem Schritt Feldbeobachtungen durchgeführt werden, um ein tiefergehendes Verständnis für das Angebot und die Aktivitäten und Tätigkeiten des pädagogischen Personals zu erhalten.
- Die Festlegung und Entwicklung des Wirkmodells erfolgt in Workshops. Durch dieses Vorgehen ist sichergestellt, dass man sich auf ein konsensuales Wirkmodell innerhalb der Einrichtung bzw. des Trägers verständigen kann.

Während der Entwicklung des Wirkmodells sollte begleitend immer wieder die Plausibilität des Modells kritisch geprüft werden, auch vor dem Hintergrund möglicher Alternativmodelle. Dies ist nötig, um einen sogenannten „confirmation bias“ (vgl. Vaessen, 2016) zu verhindern. Unter diesem Begriff wird verstanden, dass die Ergebnisse einer Wirkungsanalyse eine Programmtheorie bzw. ein Wirkmodell bestätigen, weil der Blick nur auf dieses eine Modell gerichtet wird und Alternativen nicht in Betracht gezogen werden. Durch die externe Begleitung bei der Anwendung des ‚Wirkungsradars‘ entsteht der Vorteil, dass der Moderator des Entwicklungsprozesses immer wieder auf mögliche Alternativen hinweist und so ein ‚Tunnelblick‘ vermieden wird.

4.3 Stufe 2: Effekte identifizieren

Diese Stufe des ‚Wirkungsradars‘ stellt das zentrale Element für Einrichtungen, die sich einer wirkungsorientierten Arbeitsweise verschrieben haben, dar. Ziel ist es in diesem Schritt, Veränderungen bei den Klientinnen und Klienten in den zentralen Outcome-Indikatoren zu identifizieren. Um diese Veränderungen von einer kausalen Wirkung abzugrenzen, wird bewusst von Effekten gesprochen. Durch den Begriff ‚Effekt‘ soll zudem ausgedrückt werden, dass die Veränderungen im Bezug zur durchgeführten Maßnahme bzw. Programm im Sinne von den zuvor definierten Zielen stehen.

Die Identifizierung der Effekte wird durch eine Veränderungsmessung, die im Rahmen eines wirkungsorientierten Monitorings in der jeweiligen Maßnahme bzw. Programm implementiert wird, ermöglicht. Unter einem Monitoring versteht man die „Erhebung von Daten zu aufeinander folgenden Zeitpunkten zu gleichen Merkmalen eines Programms“ (EvalWiki, 2015). In der Abgrenzung zwischen Monitoring und Evaluation zeigt sich, dass die Durchführung eines Monitorings eher Aufgabe des Programmverantwortlichen bzw. den Mitarbeitenden in einer Maßnahme ist. Externe Institute sind aber oft bei der Entwicklung von Erhebungsinstrumenten und der Auswertung der Monitoringdaten beteiligt (vgl. Klockgether, 2016, S. 179).

Im Rahmen dieser Stufe entwickelt das Institut für Praxisforschung und Evaluation zusammen mit dem Auftraggeber die Erhebungs- und Auswertungsinstrumente für die zu untersuchende Maßnahme bzw. das Programm und wertet die gewonnenen Daten aus. Bei der Implementierung der Erhebung wird darauf geachtet, dass sich das Monitoring in bereits bestehende Dokumentationsprozesse integrieren lässt, um eine zusätzliche Arbeitsbelastung der Mitarbeitenden vor Ort so gering wie möglich zu halten. Bei der Entwicklung des Erhebungsinstrumentes ist das Ziel, eine valide und reliable Erhebung der Outcome-Indikatoren zu realisieren. Daneben sollen auch die Outputs, Aktivitäten, Inputs sowie soziodemografische Daten der Klientinnen und Klienten sowie Prozess- bzw. Kontextindikatoren erfasst werden, sofern diese noch nicht über andere Dokumentationssysteme abgebildet sind.

Um die Veränderung bei den Klientinnen und Klienten aufzeigen zu können, muss eine Erhebung in dieser Stufe im Rahmen eines Längsschnittdesigns an mindestens zwei Erhebungszeitpunkten erfolgen, nämlich zu Beginn einer Maßnahme/eines Programms und am Ende bzw. beim Ausscheiden des Klienten. Es wird empfohlen zu prüfen, ob weitere Erhebungszeitpunkte innerhalb des Maßnahmeverlaufs realisiert werden können. Erfahrungen aus der Praxis zeigen, dass sich der

Aufwand durch solche weiteren Erhebungen in Grenzen hält, die Aussagen über mögliche Veränderungen aber wesentlich aussagekräftiger werden. So können durch die Integration von weiteren Messzeitpunkten (z. B. zur Mitte der Maßnahme) ein detaillierterer Verlauf aufgezeigt werden. Wurden im Wirkmodell auch Aussagen über die Persistenz und/oder den Transfer getroffen, müssen im Rahmen des Monitorings auch Nachbefragungen, also Erhebungszeitpunkte nach Beendigung der Maßnahme, vorgesehen werden.

Bei der Auswertung der erhobenen Daten werden diese, neben einer Darstellung der Gesamtveränderungen, auch detailliert nach soziodemografischen Merkmalen bzw. möglichen Prozess- bzw. Kontextmerkmalen des Programms (z. B. Gruppengröße) analysiert. Durch diesen Schritt können schon im Rahmen dieser Stufe mögliche Unterschiede in der Zielgruppe bzw. der Umsetzung herausgearbeitet und – falls nötig – konzeptionelle Weiterentwicklungen abgeleitet werden. Wird ein solches Monitoringsystem innerhalb eines Arbeitsbereiches gemeinsam mit mehreren Einrichtungen entwickelt, ist natürlich auch eine Benchmarking-Auswertung möglich, also der Vergleich der individuellen Ergebnisse einer einzelnen Einrichtung im Vergleich zu den Gesamtergebnissen aller beteiligten Einrichtungen.

Da viele Angebote der Sozialen Arbeit nicht als Einzelangebote, sondern in Gruppen stattfinden, muss bei der Datenauswertung, gerade bei der Messung von Veränderungen, auch beachtet werden, ob hierarchische Daten vorliegen und man diese Tatsache besonders in der Datenanalyse berücksichtigen muss. Betrachtet man beispielsweise die Veränderungen in einem Angebot in Kindertagesstätten, können neben den individuellen Merkmalen der Kinder (Geschlecht, Alter etc.) auch die Gruppenstruktur (Größe der Gruppe, Anzahl des Betreuungspersonals etc.) einen Einfluss auf eine mögliche Veränderung haben. Im Rahmen von Mehrebenenanalysen kann überprüft werden, ob der Einfluss der Gruppenebene so groß ist, dass die Daten mit besonderen statistischen Methoden (vgl. Hosoya, Koch & Eid, 2014) analysiert werden müssen.

Die Ergebnisse in dieser Stufe des ‚Wirkungsradars‘ geben Aufschluss darüber, ob sich darauf aufbauend weitere Analysen und Nachweise von Wirkungen lohnen. Sollten Veränderungen zuverlässig festgestellt werden, kann abschließend im Rahmen eines Workshops eine Wirkungsplausibilisierung durchgeführt werden. Hierbei werden die gefundenen Veränderungen detailliert interpretiert, um darauf aufbauend zu diskutieren, inwieweit der Untersuchungsgegenstand (Maßnahme oder Programm) für diese Veränderungen verantwortlich

ist. Mögliche Außeneinflüsse, die die Veränderungen bedingen können, werden diskutiert und zusammen mit dem Diskussionsergebnis festgehalten.

4.4 Stufe 3: Wirkung nachweisen

Der letztendlich empirische Wirkungsnachweis erfolgt im Rahmen des ‚Wirkungsradars‘ in dieser Stufe. Erst wenn diese erfolgreich durchlaufen ist und ein positives Ergebnis vorliegt, kann davon gesprochen werden, dass die Wirkung des Angebotes nachgewiesen wurde. Ziel ist es, die Wirkung kausal auf die untersuchte Maßnahme bzw. das untersuchte Programm zurückzuführen. Im Rahmen der Analyse erfolgt der Wirkungsnachweis einmalig, am besten jedoch in unterschiedlichen Kontexten. Daher unterscheidet sich dieser Analyseschritt von Stufe 2, wo versucht wurde, ein dauerhaftes Monitoringsystem zur Messung von Veränderungen aufzubauen.

Um die Wirkung kausal auf ein Angebot der Sozialen Arbeit zurückführen zu können, wird das sogenannte kontrafaktische Modell der Kausalität zugrunde gelegt. Dieses besagt, dass zur Messung eines kausalen Effektes an derselben Person zwei Messungen zum selben Zeitpunkt stattfinden müssen: Einmal, wenn die Maßnahme erfolgt ist und einmal, wenn die Maßnahme nicht erfolgt ist.

Durch die Differenz der Messwerte zwischen den Zuständen kann der kausale Effekt bestimmt werden (vgl. Morgan & Winship, 2015, S. 4). In dieser Beschreibung wird das fundamentale Problem der kausalen Inferenz (vgl. Holland, 1986) deutlich: Bei einer einzelnen Personen können nicht am selben Zeitpunkt beide Zustände gemessen werden. Entweder ist diese in der Maßnahme oder sie nimmt nicht an der Maßnahme teil. Um dieses Problem zu lösen, betrachtet man nicht die Werte einzelner Personen, sondern durchschnittliche Gruppeneffekte (vgl. Morgan & Winship, 2015, S. 46 ff.). In der Praxis werden hierzu eine Maßnahmengruppe (Personen, die an der Maßnahme teilnehmen) und eine Kontroll- bzw. Vergleichsgruppe (Personen, die nicht an der Maßnahme teilgenommen haben) gebildet.

Um eine Kontrollgruppe zu bilden, wird innerhalb der Wirkungsforschung die Durchführung von Experimenten (RTC) als „Goldstandard“ angesehen, denn: Bei einem Experiment erfolgt eine randomisierte Zuordnung der Personen zur Maßnahmen- oder zur Kontrollgruppe. D. h. allein der Zufall entscheidet, ob eine Person an der Maßnahme teilnimmt oder sich in der Kontrollgruppe befindet und daher die Maßnahme nicht in Anspruch nimmt. Würde man die Wirkung von

Angeboten der Sozialen Arbeit auf eine solche Weise untersuchen, so entstünde in vielen Fällen ein ethisches Problem: Hat man beispielsweise eine neue Maßnahme zur Unterstützung von Jugendlichen, die Probleme im Umgang mit Alkohol haben, entwickelt, sollte nicht der Zufall entscheiden, welche Jugendliche an der Maßnahme teilnehmen dürfen oder nicht. Vielmehr sollten alle Jugendlichen, die dieses Probleme haben, daran teilnehmen. Oder zumindest sollte die Teilnahme davon abhängig gemacht werden, wie ausgeprägt das Problem bei den Jugendlichen ist.

Nicht nur aus dem genannten ethischen Problemen, sondern weil ganz häufig auch praktische Schwierigkeiten mit der Bildung von randomisierten Kontrollgruppen verbunden sind, wird bei der Anwendung des ‚Wirkungsradars‘ mit so genannten quasi-experimentellen Methoden gearbeitet, die zum Ziel haben, eine Vergleichsgruppe in der Analyse zu bilden bzw. zu berücksichtigen. Hierfür gibt es mehrere statistische Methoden, die zum Einsatz kommen können (vgl. Henry, 2015). Am besten in der Praxis umsetzbar erscheint aus unserer Erfahrung der Ansatz des Matchings, genauer des ‚Propensity Scores Matchings‘ (vgl. Rosenbaum & Rubin, 1983), der deshalb auch im Rahmen des ‚Wirkungsradars‘ als Standardmethode in dieser Stufe angewendet wird.

Bei diesem Ansatz sind Maßnahmen- und Vergleichsgruppe schon vorhanden. In der Maßnahmengruppe werden alle Personen befragt, die an der Maßnahme teilnehmen. Gleichzeitig wird versucht, Personen, die nicht an der Maßnahme teilnehmen, aber ähnliche soziodemografische Merkmale haben, zu befragen und daraus eine möglichst optimale Vergleichsgruppe zu bilden. Im Rahmen des Matchings wird versucht, jedem Teilnehmenden der Maßnahmengruppe einen ‚statistischen Zwilling‘ in der Vergleichsgruppe zuzuordnen. Hierbei zieht man als Hilfsmittel im Rahmen des Propensity Score Matching den sogenannten Propensity Score heran (vgl. Caliendo & Kopeinig, 2008). Der Propensity Score ist eine Kennzahl, die für jede Person in der Maßnahmen- und Vergleichsgruppe berechnet wird. In die Berechnung fließen optimaler Weise alle zuvor im Wirkmodell definierten und in der Befragung erhobenen, Störvariablen ein, die deshalb auch kontrolliert werden sollen. Der Auswahl der Störvariablen kommt dabei eine besonders wichtige Bedeutung bei der Planung der Analyse zu, da nur durch eine gute Auswahl der Störvariablen eine hohe interne Validität und ein gutes Matching-Ergebnis erzielt wird (vgl. Cook, Pohl & Steiner, 2011; Henry, 2015, S. 153).

Nach dieser Berechnung erfolgt der Abgleich der Gruppen mit dem berechneten Propensity Score Wert. Jedem Teilnehmenden in der Maßnahmengruppe wird ein Teilnehmender aus der

Vergleichsgruppe zugeordnet, der einen möglichst gleichen bzw. ähnlichen Propensity Score Wert aufweist. Bei der Durchführung in der Praxis ist es vorteilhaft, wenn die Vergleichsgruppe größer als die Maßnahmengruppe ist, um ein möglichst optimales Matching durchführen zu können. In den meisten Fällen, wenn ein sog. 1:1-Matching durchgeführt wird, sind Maßnahme- und Vergleichsgruppe, die in die weitere Analyse einbezogen werden, gleich groß.

Ist die Bildung einer Vergleichsgruppe auf die beschriebene Art und Weise jedoch nicht möglich, so kann geprüft werden, ob mit einem sogenannten Wartekontrollgruppen-Design (vgl. Kolip & Greif, 2016) bzw. Pipeline-Ansatz (vgl. Khandker, Koolwal & Samad, 2010) gearbeitet werden kann. Vor allem bei Angeboten in Schulen und Kindertageseinrichtungen wird ein Angebot nämlich nicht in allen Klassen und Gruppen sofort eingeführt, sondern nur ein Teil startet damit, die restlichen Klassen und Gruppen beginnen (nach einem definierten Zeitraum) später. So ist dadurch einerseits sichergestellt, dass danach alle Personen das Angebot in Anspruch nehmen konnten, durch die Wartezeit andererseits aber auch eine Vergleichsgruppe gebildet. Vor dem Einsatz eines solchen Verfahrens muss jedoch geprüft werden, ob die Wartezeit aus ethischer Sicht vertretbar ist.

Wie bereits in der vorherigen Stufe ist auch beim letztendlichen *Wirkungsnachweis* – neben der Bildung einer Vergleichsgruppe – die Messung zu zwei Zeitpunkten, nämlich vor und nach der Maßnahme, notwendig. Durch diese Kombination der Vorher-/Nachher-Messung und der Bildung der Maßnahmen- und Vergleichsgruppe ist nun die notwendige Grundlage vorhanden, um den sogenannten Differenz-in-Differenz-Ansatz (vgl. Angrist & Pischke, 2009, S. 227 ff.) anzuwenden. Bei diesem Ansatz wird der kausale Effekt in einer Maßnahme wie folgt berechnet:

$$\delta_{dd} = (\bar{y}_{t2, MG} - \bar{y}_{t2, VG}) - (\bar{y}_{t1, MG} - \bar{y}_{t1, VG})$$

In dieser Berechnung werden zwei Differenzen berechnet:

- In einem ersten Schritt werden die Mittelwerte für beide Gruppen bei der Messung nach der Maßnahme (t2) subtrahiert. Das Gleiche geschieht mit den Mittelwerten der Gruppen bei der Messung vor der Maßnahme (t1). Durch diese Berechnung werden u. a. unterschiedliche Ausgangswerte zwischen den Gruppen am Anfang der Maßnahme berücksichtigt.

- Im zweiten Schritt werden dann diese beiden Differenzen (Differenz der Gruppen bei der Nachher-Messung und Differenz der Gruppen bei der Vorher-Messung) voneinander abgezogen.

Dieses Ergebnis stellt dann den kausalen Effekt für das untersuchte Merkmal dar. Nachfolgende Grafik veranschaulicht die Berechnung nochmals:

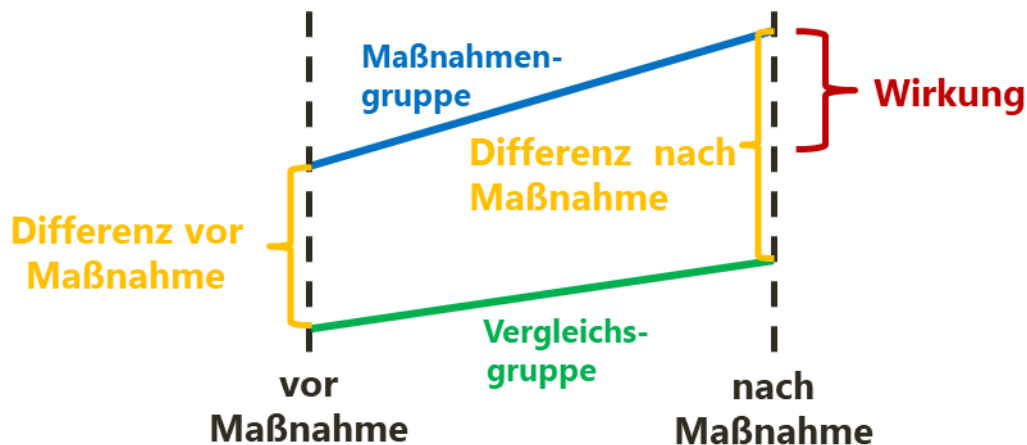


Abbildung 10: Differenz-in-Differenz Ansatz (Darstellung in Anlehnung an Caspari, 2009, S. 194)

Durch den Einbezug der Störvariablen in das Matching ist sichergestellt, dass sich die Personen in den berücksichtigten Merkmalen sehr ähnlich sind. Bei den nun gefundenen Effekten kann man daher nun tatsächlich die Aussage treffen, dass die kontrollierten Störvariablen keinen Einfluss auf die Veränderung haben, dafür aber die untersuchte Maßnahme. Durch eine optimale Umsetzung der beschriebenen Methoden und guter Datenlage führt das dargestellte quasi-experimentelle Vorgehen „zu ähnlich robusten Ergebnissen über die Wirkungen einer Maßnahme (...) wie ein experimentelles Design“ (Caspari, 2009, S. 198).

Abschließend ist anzumerken, dass es bei Angeboten in der Sozialen Arbeit häufig sehr schwierig, oft auch gar nicht möglich ist, alle möglichen Störvariablen zu kontrollieren. Dies ist angesichts komplexer sozialer Wirklichkeiten und vielfältiger Wechselwirkungen auch evident. Trotzdem sollte bei dem Versuch, einen Wirkungsnachweis zu erbringen, der Anspruch vorhanden sein, alle denkbaren Störvariablen so gut wie möglich abzubilden und auch zu berücksichtigen, um eine möglichst hohe interne Validität zu erzielen. So ist dann auch sichergestellt, dass die gefundenen Effekte mit einer sehr hohen Wahrscheinlichkeit auch als Wirkungen auf das untersuchte Angebot zurückgeführt werden können.

4.5 Stufe 4: Wirkmechanismen analysieren

Im Hinblick auf eine wirkungsorientiertere Ausgestaltung von Angeboten der Sozialen Arbeit und den damit verbundenen Aufbau evidenzbasierten Wissens für die fachliche Arbeit kommt dieser Stufe eine besonders hohe Bedeutung zu. Denn: Bei diesem Analyseschritt geht es nun darum, die Black-Box des Wirkungsnachweises zu öffnen und im Sinne einer sogenannten White-Box-Evaluation nicht nur zu fragen, *ob* etwas wirkt, sondern auch, *wie* die Wirkung entstanden sein könnte.

Im Rahmen der Analyse von Wirkmechanismen spielt das am Anfang der Untersuchung aufgestellte theoretische Wirkmodell erneut eine entscheidende Rolle. In diesem Modell werden mögliche Mechanismen festgelegt, von denen angenommen werden kann, dass sie bei der Durchführung der Maßnahme die gewünschte Wirkung entfalten. Es gilt nun – im Rahmen eines sogenannten Pattern Matching (vgl. Giel, 2013, S. 227 ff.) – abzugleichen, ob die zuvor theoretisch und vor dem Hintergrund der Kenntnisse und Erfahrungen der Fachkräfte definierten Mechanismen auch empirisch nachgewiesen werden können.

Um diesen empirischen Nachweis erbringen zu können, stehen unterschiedliche quantitative und vor allem auch qualitative Forschungsmethoden zu Verfügung. In einem ersten Schritt werden – im Rahmen dieser Stufe des ‚Wirkungsradars‘ – die gefundenen Wirkungen tiefergehend in Hinblick auf verschiedene Zielgruppen untersucht. Es werden Unterschiede zwischen den Teilgruppen berechnet und geprüft, ob in allen Teilgruppen die gleichen Effekte auftreten oder ob Gruppen vorhanden sind, die von der Maßnahme nicht so stark profitieren. Dies ist auch dann möglich, wenn in Stufe 2 nur eine Veränderungsmessung vorgenommen wurde, denn auch eine tiefergehende Analyse der gefundenen Effekte bei den Klientinnen und Klienten kann sehr interessante Aspekte und Perspektiven für die konzeptionelle Debatte und Weiterentwicklung der untersuchten Maßnahmen erbringen.

Im Rahmen einer umfassenden quantitativen Analyse von Wirkmechanismen kann die Methode des Strukturgleichungsmodells eingesetzt werden. Im Rahmen von Strukturgleichungsmodellen können theoretisch generierte Hypothesensysteme empirisch überprüft werden. Vorteil der Methode ist, dass man mehrere abhängige Variablen definieren kann und auch latente Konstrukte, also Sachverhalte und Phänomene, die nicht direkt messbar sind, Berücksichtigung finden (vgl. Arzheimer, 2016). Sofern die einzelnen Komponenten des definierten Wirkmodells valide und

reliabel operationalisiert werden konnten und eine Erhebung der Indikatoren durchgeführt wurde, kann die Analyse mit einem Strukturgleichungsmodell erfolgen. Allerdings kann mit diesem quantitativen Ansatz nur ein theoretisches Wirkmodell empirisch überprüft werden. Der Ansatz liefert keine weiteren Erklärungen für die entstandenen Wirkungen und es besteht – wie bereits unter Abschnitt 4.2 aufgeführt – die Gefahr eines sogenannten ‚confirmation bias‘ (vgl. Vaessen, 2016). Dieser sagt aus, dass durch die Evaluationsergebnisse ein bestimmtes Wirkmodell angenommen wird, obwohl es auch alternative Erklärungsansätze gibt, um die Entstehung der Wirkung zu erklären. Insofern ist zu empfehlen, den Einsatz von Strukturgleichungsmodelle mit qualitativen Methoden zu flankieren.

Im Rahmen der qualitativen Methoden haben sich in der Praxis Fokusgruppen und Experteninterviews als praktikabel erwiesen. Vorteil dieser Methoden ist es, dass man hier noch einmal umfassender die Sicht der Klientinnen und Klienten auf die untersuchte Leistung erfassen und damit auch in die Wirkungsanalyse einbeziehen kann. Bei der Anwendung solcher Methoden empfiehlt es sich, so viele Stakeholder wie möglich einzubinden. Neben den Mitarbeitenden und den Teilnehmenden der Maßnahme könnten dies auch Leitungskräfte der Einrichtung und Kooperationspartner sein. Ziel ist es, eine möglichst vielschichtige Sicht auf die Maßnahme zu erhalten und im Rahmen der Fokusgruppen und Experteninterviews mögliche Mechanismen herauszuarbeiten. In der abschließenden Analyse des Datenmaterials kann überprüft werden, an welchen Stellen ähnliche Mechanismen identifiziert werden konnten, aber auch wo es unterschiedliche Erklärungs- und Bewertungsansätze gibt.

Eine weitere Methode, die im Rahmen dieser Analysestufe eingesetzt werden kann, ist das Process Tracing. Diese Methode kommt aus den Politikwissenschaften (vgl. Munro, 2016), wird aber auch in Evaluationsstudien eingesetzt (vgl. u.a. Punton & Welle, 2015; Schmitt & Beach, 2015). Hier wird mit Einzelfällen im Rahmen einer Fallstudie gearbeitet. Es wird versucht, den kausalen Prozess zwischen der Maßnahme bzw. dem Programm und dem Outcome zu identifizieren. Durch dieses Vorgehen können bestehende Annahmen über Wirkmechanismen bestätigt, aber auch alternative Erklärungen entwickelt werden. Mit Hilfe verschiedener Tests können klare Aussagen generiert werden, ob eine Bestätigung dieser Mechanismen möglich ist oder nicht (vgl. Munro, 2016, S. 84).

Welche der soeben beschriebenen Methoden bei der Anwendung des ‚Wirkungsradars‘ eingesetzt werden, hängt vom Untersuchungsgegenstand, aber auch, mit Blick auf den zu betreibenden Aufwand, von den vorhandenen Ressourcen ab. Es können einzelne Methode eingesetzt werden, aber auch eine Triangulation von verschiedenen der beschriebenen Methoden ist möglich, um dadurch umso aussagekräftigere Ergebnisse über die zugrundeliegenden Mechanismen erhalten zu können.

4.6 Stufe 5: Effizienz belegen

In der letzten Stufe des ‚Wirkungsradars‘ wird die Effizienz des untersuchten Angebots belegt – sofern dies möglich ist. Im Bereich der Evaluation wird vor allem im Gesundheitsbereich immer auch die Effizienz von Maßnahmen im Rahmen von gesundheitsökonomischen Evaluationen betrachtet (vgl. u. a. Scherenberg, 2018; Schöffski & Schulenburg, 2012). In den Arbeitsfeldern der Sozialen Arbeit hat sich für Effizienzbetrachtungen in den letzten Jahren der Ansatz des Social Return on Investment etabliert (vgl. Schober & Then, 2015; The SROI Network, 2012). Bei diesem Ansatz wird versucht, die gefundenen Wirkungen eines Angebotes zu monetarisieren. Dies kann beispielsweise durch die Berechnung von Alternativkosten geschehen, die angefallen wären, wenn die Maßnahme nicht die gezeigte Wirkung erzielt hätte. Den monetarisierten Wirkungen werden dann die finanziellen Mittel, die für das Angebot zur Verfügung gestellt wurden (finanzieller Input), gegenübergestellt. Das errechnete Verhältnis wird als Social Return on Investment bezeichnet.

Der hier beschriebene SROI-Ansatz wird auch im Rahmen des ‚Wirkungsradars‘ eingesetzt, um die Effizienz einer Maßnahme zu belegen. Um eine belastbare und valide Bewertung der Effizienz durchführen zu können, werden im Rahmen der Berechnungen nur nachvollziehbare Monetarisierungen betrachtet, d. h. es werden nur solche aufgenommen, die naheliegend und zum Zeitpunkt der Analyse belegbar sind. Auch wird auf mögliche finanzielle Effekte, die in einer sehr großen Zukunft (länger als 3–5 Jahre) liegen, verzichtet. Abhängig davon, welche Wirkungen betrachtet werden, also solche auf der Ebene der Klientinnen und Klienten oder die gesellschaftliche Wirkung insgesamt, werden auch nur diese Wirkungen monetarisiert. Nicht betrachtet im Rahmen dieser Stufe werden finanzielle ‚Wirkungen‘, die aufgrund der Mitarbeitenden bzw. Einkäufe der sozialen Einrichtung erfolgen (z. B. Steuern). Die Herausnahme

dieser Effekte scheint auch angebracht, da diese keine direkte Aussagekraft über die Wirksamkeit eines Angebotes der Sozialen Arbeit in fachlicher Hinsicht haben.

4.7 Abschluss: Ergebnisse berichten

Bei der Berichtslegung der Ergebnisse einer Untersuchung, in der der ‚Wirkungsradar‘ zur Anwendung kommt, wird der Social Reporting Standard (vgl. Social Reporting Initiative e.V., 2014) angewendet. Der Social Reporting Standard stellt eine standardisierte Berichterstattung für soziale Angebote dar. Neben Informationen über die Organisation und ihre Vision werden die Angebote detailliert dargestellt. In dieser Darstellung wird auch auf eine mögliche Wirkung der Angebote eingegangen. Es werden allerdings keine Vorgaben über den Nachweis der Wirkungen gemacht, so dass hier der ‚Wirkungsradar‘ als ein mögliches Instrument zum Einsatz kommen kann.

Neben der Berichtlegung mit dem Social Reporting Standard erstellt das Institut für Praxisforschung und Evaluation – sofern nötig – einen ergänzenden Ergebnisband. Weiterhin werden in einem Methodenbericht die verschiedenen Schritte der Analyse und die eingesetzten Methoden nachvollziehbar beschrieben, damit diese gegenüber Stakeholdern offengelegt werden können.

Neben den beschriebenen Berichtsformaten werden auf Wunsch weitere Auswertungsformate zur Verfügung gestellt: Für die konzeptionelle Weiterarbeit werden die Ergebnisse der Analyse in Form eines Dashboards aufbereitet. Hierdurch können die Einrichtungen die gewonnenen Daten selbstständig interaktiv tiefergehend analysieren und zum Beispiel Gruppenunterschiede genauer betrachten. Auch ist eine Aktualisierung des Dashboards um neu erhobene Daten, beispielsweise durch das in Stufe 2 des ‚Wirkungsraders‘ etablierte Monitoring, jederzeit möglich.

Um die gefundenen Ergebnisse und zentralen Wirkmechanismen in die Praxis transferieren zu können, ist eine verständliche und einfache Beschreibung dieser oft komplexen Zusammenhänge nötig (vgl. Macsenare, 2018, S. 31). Daher werden auf Wunsch die zentralen Befunde und Wirkmechanismen besonders mit Blick auf die Praxis aufbereitet, beispielsweise im Rahmen von Infografiken oder Informationsflyern.

5. Grenzen der Wirkungsanalyse in der Sozialen Arbeit

Nachdem der ‚Wirkungsradar‘ als eine Toolbox für die Wirkungsanalyse in den Feldern der Sozialen Arbeit vorgestellt wurde, soll nun am Schluss noch auf mögliche Grenzen der Wirkungsanalyse eingegangen werden.

Eine Grenze ergibt sich, wenn bestimmte methodische Anforderungen an den kausalen Wirkungsnachweis nicht erfüllt sind. Kann keine Vergleichsgruppe befragt werden und/oder keine Messungen vor und nach der Maßnahme durchgeführt werden, so ist logischerweise auch eine kausale Rückführung von gefundenen Veränderungen auf die Maßnahme nicht möglich. Konsequenterweise muss in diesem Fall dann auch klar formuliert werden, dass ein empirischer Wirkungsnachweis nicht erfolgen kann und eben ‚nur‘ eine Wirkungsplausibilisierung möglich ist. Gelangt man während der Untersuchung an einen solchen Punkt, so muss dieses Ergebnis transparent kommuniziert werden. Zudem erscheint es gegenüber dem Kostenträger und der Politik sinnvoll, dann auch aufzuzeigen, warum kein kausaler Nachweis der Wirksamkeit möglich ist.

Bei der Durchführung von Wirkungsanalysen in der Praxis zeigt sich häufig auch das Problem, dass eine Vorher-Messung nicht möglich ist, da die Untersuchung zu spät beauftragt wurde. Hier empfiehlt es sich, dass Einrichtungen und Träger möglichst vorausschauend und frühzeitig eine entsprechende Untersuchung planen und gegebenenfalls vergeben. Umso früher eine Wirkungsanalyse beauftragt wird, desto besser kann diese auch in die alltägliche Arbeit integriert und damit der Aufwand für die Mitarbeitenden vor Ort minimiert werden.

Doch auch wenn ein experimentelles oder quasi-experimentelles Design mit einer Vergleichsgruppe umgesetzt werden kann, kann es zu Einschränkungen auf die Aussagekraft und die Belastbarkeit der Befunde kommen. Diese Einschränkungen stehen sehr häufig in Zusammenhang mit der internen und externen Validität der Untersuchung. Interne Validität bedeutet, „dass ein hypothesenbasiertes Evaluationsdesign in der Lage ist, eine eindeutige Aussage über die Annahme oder die Ablehnung der entsprechenden Hypothesen zu liefern“ (Gollwitzer & Jäger, 2014, S. 42). In Bezug auf Wirkungsuntersuchungen bedeutet dies, dass eine gefundene Wirkung nur bei Vorliegen einer hohen internen Validität kausal auf die untersuchte Maßnahme zurückzuführen ist. Demgegenüber steht die externe Validität, die darlegt, inwiefern es möglich ist, „die Ergebnisse der Evaluationsuntersuchung auf andere Kontexte zu generalisieren.“

(Gollwitzer & Jäger, 2014, S. 42). Im Hinblick auf die Übertragbarkeit der gefundenen Ergebnisse sollte auch die externe Validität möglichst hoch sein. In der Praxis zeigt sich nun allerdings, dass es nie möglich ist, beide Validitäten im Rahmen eines Forschungsdesigns gleichzeitig vollständig zu erfüllen. Am einfachsten ist dies an Laborexperimenten zu beschreiben: Dadurch, dass in einem Labor Störfaktoren komplett eliminiert werden können, ist die interne Validität sehr hoch. Allerdings ist gleichzeitig die externe Validität gering, da es schwierig ist die im Labor gefundenen Ergebnisse komplett in die Realität zu übertragen.

Bei der praktischen Durchführung von experimentellen oder quasi-experimentellen Designs gibt es eine Vielzahl an Problemen, die einen Einfluss auf die interne und externe Validität der Untersuchung haben und damit auch eine quasi ‚natürliche Grenze‘ für Wirkungsanalysen darstellen:

Im Hinblick auf die interne Validität kann es zu Verzerrungen durch die Befragten kommen. Sollen Veränderungen auf der Ebene der Klientinnen und Klienten erhoben werden, gibt es die Möglichkeit der Selbst- und Fremdeinschätzung. Dies bedeutet, dass entweder die Klientinnen und Klienten direkt befragt werden, oder bei einer Fremdeinschätzung z. B. die Mitarbeitenden eine Beurteilung aus ihrer Sicht abgeben.

Bei einer Bewertung durch die Klientinnen und Klienten selbst kann es zu Verzerrungseffekten durch die sogenannte soziale Erwünschtheit kommen. Dies bedeutet, dass die Befragten besonders positiv oder negativ antworten, um sich besonders gut darzustellen (vgl. Gollwitzer & Jäger, 2014, S. 67 f.). Gerade im Hinblick auf Nutzerinnen und Nutzer der Angebote der Sozialen Arbeit, die wegen verschiedenen Problemlagen diese Angebote oft nutzen, ist die Gefahr groß, dass bei einer Selbsteinschätzung positivere Bewertungen abgegeben werden als dies der Realität entspricht. Dieses Phänomen bestätigt sich bei der Durchführung vieler bisheriger Befragungen.

Ein weiterer Aspekt, der zu Verzerrungen in Selbsteinschätzungen führen kann, ist, dass Klientinnen und Klienten eine Maßnahme und das Hilfsangebot aus Dankbarkeit zu positiv bewerten (vgl. Scheirer, 1978). Auch diesen Aspekt sollte man schon im Vorfeld der Erhebung bei der Entwicklung von Erhebungsinstrumenten beachten. Bei der Interpretation der Ergebnisse gilt es, diese kritisch zu hinterfragen und vor diesem Hintergrund zu diskutieren, ob solche Verzerrungseffekte vorliegen und dadurch die Reichweite und Belastbarkeit der Befunde einschränken.

Aber auch Fremdeinschätzungen sind nicht gefeit vor Verzerrungen. So kann es durchaus vorkommen, dass pädagogische Fachkräfte Klientinnen und Klienten bewusst (oder unbewusst) schlechter einschätzen als es der Realität entspricht, um ihre Arbeit und das Angebot zu

legitimieren. Hier ist es wichtig, den Mitarbeitenden am Anfang der Untersuchung klar zu kommunizieren, welche Ziele mit der Untersuchung verbunden sind, um den Mitarbeitenden Befürchtungen und Ängste vor eventuellen negativen Auswirkungen (Rationalisierung, Kürzungen usw.) auf die eigene Arbeit zu nehmen.

An dieser Stelle können nur einige wenige Probleme im Hinblick auf die interne Validität beschrieben werden. Umfangreichere Ausführungen sind u. a. bei Müller (2017) sowie Shadish et al. (2002) zu finden. Im Rahmen einer Wirkungsanalyse muss eine realistische Einschätzung der internen und externen Validität dringend erfolgen und auch transparent kommuniziert werden. Im Hinblick auf die externe Validität, also die Generalisierbarkeit der Ergebnisse, wird bei der Anwendung des ‚Wirkungsradars‘ durch den Denkansatz der Realistic Evaluation nur der Anspruch auf eine Theorie mittlerer Reichweite erhoben, die durch die Untersuchung bewiesen werden kann. Ganz bewusst ist daher von vornherein keine Generalisierung der gefundenen Wirkungen impliziert, da diese immer vom praktischen Kontext, in dem ein Angebot stattfindet, abhängt.

Grundsätzlich gilt: Bei der Umsetzung von Wirkungsanalysen gibt es keine Patentrezepte zur Erzielung einer möglichst hohen internen und externen Validität. Im Hinblick auf die interne Validität empfiehlt es sich aber immer – falls möglich – verschiedene Befragungsgruppen als Datenquellen im Rahmen einer Datentriangulation in die Wirkungsanalyse einzubinden. Veränderungen können beispielsweise durch die Erhebung im Rahmen einer Selbsteinschätzung durch die Klientinnen und Klienten und zusätzlich im Rahmen einer Fremdeinschätzung durch Fachkräfte erhoben werden. Die Ergebnisse der Einschätzungen können abgeglichen werden und man erhält somit eine realistischere Bewertung.

Um die externe Validität einer Wirkungsanalyse zu erhöhen, muss die Wirkung des untersuchten Angebotes in unterschiedlichen Kontexten überprüft werden. Unter anderem fordert auch die Grüne Liste Prävention, dass eine Maßnahme mindestens in zwei unterschiedlichen Kontexten wirksam war, bevor diese in die Liste aufgenommen wird (vgl. Gollwitzer & Jäger, 2014, S. 120 f.).

Eine weitere Methode um die Wirkung eines Angebotes zu generalisieren, sind Metastudien (vgl. Eisend, 2014). Im Rahmen einer Metaanalyse werden verschiedene Untersuchungen zum zuvor definierten Untersuchungsgegenstand einbezogen und die Ergebnisse der Untersuchungen miteinander verglichen. Durch dieses Vorgehen können Effekte identifiziert werden, die sich in allen eingeschlossenen Untersuchungen zeigen, aber auch Unterschiede bei Ergebnissen erklärt

werden. Gerade in den Arbeitsfeldern der Sozialen Arbeit wäre es daher, auch in Deutschland, wünschenswert, wenn Metaanalysen zu verschiedenen Angeboten stärker gefördert und durchgeführt würden, wie dies z. B. die Campbell Collaboration (vgl. Campbell Collaboration, o.J.) vollzieht. Ein weiterer Aspekt, der in diesem Zusammenhang mitgedacht werden sollte, ist die Beauftragung von Wirkungsanalysen innerhalb eines Arbeitsfeldes über eine einzelne Einrichtung hinaus. So können Einrichtungen und Träger mit einem ähnlichen Angebot gemeinsam eine Wirkungsanalyse in Auftrag geben. Neben der Möglichkeit des Benchmarkings sind dann auch validere Aussagen über die Generalisierbarkeit der gefundenen Wirkungen möglich. Dies ist besonders dann der Fall, wenn sich Einrichtungen aus verschiedenen Bundesländern und von verschiedenen Trägern zusammentun.

Trotz allem stellen die hier dargestellten Grenzen der Wirkungsanalyse kein Argument dar, sich nicht mit dem Thema Wirkungsorientierung und Wirkungsanalyse in der Sozialen Arbeit zu beschäftigen. Vielmehr ist ein differenzierter und realistischer, an der jeweiligen Praxis vor Ort orientierter Umgang damit nötig! Der vorgestellte ‚Wirkungsradar‘ als empirische Toolbox bietet genau diesen Zugang für die verschiedenen Arbeitsfelder der Sozialen Arbeit.

6. Literaturverzeichnis

- Albus, S., Greschke, H., Klingler, B., Messmer, H., Michael, H.-G., Otto, H.-U. et al. (2010). Abschlussbericht der Evaluation des Bundesmodellprogramms „Qualifizierung der Hilfen zur Erziehung durch wirkungsorientierte Ausgestaltung der Leistungs-, Entgelt- und Qualitätsvereinbarungen nach §§ 78a ff SGB VIII“ (Wirkungsorientierte Jugendhilfe). Münster: Waxmann.
- Angrist, J. D. & Pischke, J.-S. (2009). Mostly harmless econometrics: an empiricist's companion. Princeton: Princeton University Press.
- Arzheimer, K. (2016). Strukturgleichungsmodelle: eine anwendungsorientierte Einführung (Methoden der Politikwissenschaft). Wiesbaden: Springer VS.
- Balzer, L. (2012). Der Wirkungsbegriff in der Evaluation – eine besondere Herausforderung. In G. Niedermair (Hrsg.), Evaluation als Herausforderung der Berufsbildung und Personalentwicklung (1. Auflage, S. 125–141). Linz: Trauner.
- Bamberger, M., Vaessen, J. & Raimondo, E. (2016). Dealing with complexity in development evaluation: a practical approach. Thousand Oaks: SAGE Publications, Inc.
- Bartsch, S., Beywl, W. & Niestroj, M. (2016). Der Programmbaum als Evaluationsinstrument. In S. Giel, K. Klockgether & S. Mäder (Hrsg.), Evaluationspraxis: Professionalisierung – Ansätze – Methoden (2. Auflage, S. 89–111). Münster: Waxmann Verlag.
- Beywl, W. & Niestroj, M. (2009). Der Programmbaum. Landmarke wirkungsorientierter Evaluation. In W. Beywl & M. Niestroj (Hrsg.), Das ABC der wirkungsorientierten Evaluation: Glossar – deutsch/englisch – der wirkungsorientierten Evaluation (2. Auflage, S. 137–149). Köln: Univision – Inst. für Evaluation Dr. Beywl und Associates.
- Bibliographisches Institut. (o. J.). Wirkung, die. Zugriff am 10.5.2018. Verfügbar unter: <https://www.duden.de/rechtschreibung/Wirkung>
- Bleck, C. (2016). Qualität, Wirkung oder Nutzen: Zentrale Zugänge zu Resultaten Sozialer Arbeit in professionsbezogener Reflexion. In S. Borrmann & B. Thiessen (Hrsg.), Wirkungen Sozialer Arbeit: Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin (1. Auflage, S. 107–124). Opladen, Berlin & Toronto: Verlag Barbara Budrich.
- Borrmann, S. & Thiessen, B. (Hrsg.). (2016). Wirkungen Sozialer Arbeit: Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin (1. Auflage). Opladen, Berlin & Toronto: Verlag Barbara Budrich.
- Caliendo, M. & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys, 22 (1), 31–72. doi:10.1111/j.1467-6419.2007.00527.x
- Campbell Collaboration. (o. J.). Our Vision, Mission and Key Principles. Campbell Collaboration. Zugriff am 24.7.2018. Verfügbar unter: <https://campbellcollaboration.org/about-campbell/vision-mission-and-principle.html>
- Caspari, A. (2009). „Rigore“ Wirkungsevaluation – methodische und konzeptionelle Ansätze der Wirkungsmessung in der Entwicklungszusammenarbeit. Zeitschrift für Evaluation, 8 (2), 182–213.

- Clark, C., Rosenzweig, W., Long, D. & Olsen, S. (2004). Double Bottom Line Project Report – Assessing Social Impact in Double Bottom Line Ventures – Methods Catalog. No. 13. University of California Berkeley, Center for Responsible Business. Zugriff am 17.5.2018. Verfügbar unter: <https://escholarship.org/uc/item/80n4f1mf>
- Cook, T. D., Pohl, S. & Steiner, P. M. (2011). Die relative Bedeutung der Kovariatenwahl, Reliabilität und Art der Datenanalyse zur Schätzung kausaler Effekte aus Beobachtungsdaten. *Zeitschrift für Evaluation*, 10 (2), 203–224.
- DeGEval – Gesellschaft für Evaluation e. V. (Hrsg.). (2017). DeGEval – Gesellschaft für Evaluation Standards für Evaluation. Erste Revision 2016. DeGEval – Gesellschaft für Evaluation e. V. Zugriff am 1.6.2018. Verfügbar unter: https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf
- Eisend, M. (2014). Metaanalyse (Sozialwissenschaftliche Forschungsmethoden) (1. Auflage., Band 8). München Mering: Rainer Hampp Verlag.
- Eppler, N., Miethe, I. & Schneider, A. (2011). Qualitative und quantitative Wirkungsforschung: Ansätze, Beispiele, Perspektiven (1. Auflage). Opladen: Verlag Barbara Budrich.
- EvalWiki. (2015). Monitoring (21. Februar 2015). Zugriff am 7.6.2018. Verfügbar unter: https://eval-wiki.org/w_glossar/index.php?title=Monitoring&toldid=1363
- Fulbright-Anderson, K., Kubisch, A. & Connell, J. (1998). New Approaches to Evaluating Community Initiatives, vol. 2, Theory, Measurement, and Analysis. Washington, D.C.: Aspen Institute.
- Giel, S. (2013). Theoriebasierte Evaluation: Konzepte und methodische Umsetzungen. Münster: Waxmann.
- Gollwitzer, M. & Jäger, R. S. (2014). Evaluation kompakt (2. Auflage). Weinheim Basel: Beltz.
- Grinnell, R. M., Gabor, P. & Unrau, Y. A. (2016). Program evaluation for social workers: foundations of evidence-based programs (7th Edition.). Oxford ; New York: Oxford University Press.
- Harries, E., Hodgson, L. & Noble, J. (2014). Creating your Theory of Change. NPC's practical guide. NPC New Philanthropy Capital. Zugriff am 28.8.2017. Verfügbar unter: <https://www.thinknpc.org/publications/creating-your-theory-of-change/creating-your-theory-of-change-3/?post-parent=12372>
- Haunberger, S. & Baumgartner, E. (2017). Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. *Zeitschrift für Evaluation*, 16 (1), 121–145.
- Henry, G. T. (2015). Comparison Group Designs. In K.E. Newcomer, H.P. Hatry & J.S. Wholey (Hrsg.), *Handbook of Practical Program Evaluation* (4. Auflage, S. 137–157). New Jersey: Jossey-Bass.
- Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81 (396), 954–960.
- Horcher, G. (2013). New Public Management. In K. Grunwald, G. Horcher & B. Maelicke (Hrsg.), *Lexikon der Sozialwirtschaft* (2. Auflage, S. 693–694). Baden-Baden: Nomos Verlagsgesellschaft.
- Horcher, G. & Zängl, P. (2013). Outcome. In K. Grunwald, G. Horcher & B. Maelicke (Hrsg.), *Lexikon*

- der Sozialwirtschaft (2. Auflage, S. 735). Baden-Baden: Nomos Verlagsgesellschaft.
- Hosoya, G., Koch, T. & Eid, M. (2014). Längsschnittdaten und Mehrebenenanalyse. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66 (1), 189–218. doi:10.1007/s11577-014-0262-9
- Kehl, K., Glänzel, G., Then, V. & Mildenerberger, G. (2016). CSI-Transparenzgutachten: Möglichkeiten, Wirkungen (in) der Freien Wohlfahrtspflege zu messen. Zugriff am 10.5.2017. Verfügbar unter: http://www.bagfw.de/uploads/media/CSI_Transparenzgutachten_2016.pdf
- Khandker, S. R., Koolwal, G. B. & Samad, H. A. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington: The World Bank.
- Klockgether, K. (2016). Der Umgang mit Monitoring in der Evaluation. In S. Giel, K. Klockgether & S. Mäder (Hrsg.), *Evaluationspraxis. Professionalisierung – Ansätze – Methoden*. (2. Auflage, S. 177–197). Münster: Waxmann Verlag.
- Knowlton, L. W. & Phillips, C. C. (2013). *The logic model guidebook: better strategies for great results* (2. Auflage). Los Angeles: SAGE.
- Kolip, P. & Greif, N. (2016). *Evaluation Programm Klasse 2000: Zusammenfassender Abschlussbericht*. Zugriff am 5.6.2018. Verfügbar unter: http://www.klasse2000.de/fileadmin/user_upload/Abschlussbericht-64Seiten.pdf
- Kratz, D. (2017). Wie Wirkungsorientierung die Adressat_innen vergisst. *Sozial Extra*, 41 (4), 32–34. doi:10.1007/s12054-017-0057-0
- Landespräventionsrat Niedersachsen. (2011). *Grüne Liste Prävention. Auswahl- und Bewertungskriterien für die CTC Programm-Datenbank*. Zugriff am 2.6.2018. Verfügbar unter: https://www.gruene-liste-praevention.de/communities-that-care/Media/Grne_Liste_Bewertungskriterien.pdf
- Macsenare, M. (2018). Hilfen zur Erziehung – Praxis trifft Forschung, Forschung trifft Praxis: Was wir voneinander wissen (wollen). *Forum Jugendhilfe*, (1), 28–32.
- Markgraf, D. (2018). Qualität. *Gabler Wirtschaftslexikon*. Zugriff am 14.5.2018. Verfügbar unter: <https://wirtschaftslexikon.gabler.de/definition/qualitaet-45908/version-269195>
- Merchel, J. (2013). *Qualitätsmanagement in der Sozialen Arbeit.: Eine Einführung*. (4. Auflage). Weinheim: Beltz Juventa Verlag.
- Merton, R. K. (1968). *Social Theory and Social Structure*. New York: The Free Press.
- Morgan, S. L. & Winship, C. (2015). *Counterfactuals and causal inference: methods and principles for social research (Analytical methods for social research)* (2. Auflage). New York, NY: Cambridge University Press.
- Müller, C. E. (2017). Kausale Wirkungsevaluation zwischen methodischen Anspruch und empirischer Praxis. In R. Stockmann & W. Meyer (Hrsg.), *Die Zukunft der Evaluation: Trends, Herausforderungen, Perspektiven* (1. Auflage, S. 205–222). Münster: Waxmann Verlag GmbH.
- Muno, W. (2016). Fallstudien und Process Tracing in der Vergleichenden Politikwissenschaft. In H.-J. Lauth, M. Kneuer & G. Pickel (Hrsg.), *Handbuch Vergleichende Politikwissenschaft* (S. 79–90). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-02338-6_6

- Newcomer, K. E., Hatry, H. P. & Wholey, J. S. (2015). Planning and Designing Useful Evaluations. In K.E. Newcomer, H.P. Hatry & J.S. Wholey (Hrsg.), Handbook of Practical Program Evaluation (4. Auflage). New Jersey: Jossey-Bass.
- Otto, H.-U. (2007). Zum aktuellen Diskurs um Ergebnisse und Wirkungen im Feld der Sozialpädagogik und Sozialarbeit – Literaturvergleich nationaler und internationaler Diskussion. Berlin: Arbeitsgemeinschaft für Kinder- und Jugendhilfe – AGJ.
- Otto, H.-U., Polutta, A. & Ziegler, H. (2007). What Works – Welches Wissen braucht die Soziale Arbeit?: Zum Konzept evidenzbasierter Praxis (1. Auflage). Opladen: Budrich.
- Pawson, R. (2010). Middle Range Theory and Program Theory Evaluation: From Provenance to Practice. In J. Vaessen & F.L. Leeuw (Hrsg.), Mind the Gap. Perspectives on Policy Evaluation and the Social Sciences. (S. 171–202). New Brunswick, London: Transaction Publishers.
- Pawson, R. & Tilley, N. (1997). Realistic evaluation. London ; Thousand Oaks, Calif: Sage.
- PHINEO gAG. (2014). Engagement mit Wirkung: Die Phineo-Analysemethode im Detail. Verfügbar unter: http://www.phineo.org/downloads/PHINEO_Engagement-mit-Wirkung.pdf
- PHINEO gAG. (2018). Gelsenkirchen wirkt! Zugriff am 12.5.2018. Verfügbar unter: <https://www.phineo.org/news/neues-aus-der-zivilgesellschaft/gelsenkirchen-wirkt-2018-01-11/back-245>
- Polutta, A. (2013). Wirkungsorientierung. In K. Grunwald, G. Horcher & B. Maelicke (Hrsg.), Lexikon der Sozialwirtschaft (2. Auflage, S. 1108–1109). Baden-Baden: Nomos Verlagsgesellschaft.
- Punton, M. & Welle, K. (2015). Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation? (CDI Practice Paper). IDS Institute of Development Studies. Zugriff am 23.6.2018. Verfügbar unter: http://opendocs.ids.ac.uk/opendocs/bitstream/123456789/5997/1/CDIPracticePaper_10.pdf
- Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Hrsg.), Handbuch der sozialwissenschaftlichen Datenanalyse (S. 239–258). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92038-2_11
- Rauscher, O., Mildenerger, G. & Krlev, G. (2015). Wie werden Wirkungen identifiziert? Das Wirkungsmodell. In C. Schober & V. Then (Hrsg.), Praxishandbuch Social Return on Investment: Wirkung sozialer Investitionen messen (S. 41–57). Stuttgart: Schäffer-Poeschel Verlag.
- Renner, I. (2012). Wirkungsevaluation „Keiner fällt durchs Netz“: ein Modellprojekt des Nationalen Zentrums Frühe Hilfen. (Nationales Zentrum Frühe Hilfen (NZFH), Hrsg.). Köln: BZgA Bundeszentrale für gesundheitliche Aufklärung. Zugriff am 28.6.2018. Verfügbar unter: https://www.fruehehilfen.de/fileadmin/user_upload/fruehehilfen.de/downloads/Wirkungsevaluation.pdf
- Rosenbaum, P. R. & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70 (1), 41–55.
- Schaarschuch, A. & Oelerich, G. (2005). Theoretische Grundlagen und Perspektiven sozialpädagogischer Nutzerforschung. In G. Oelerich & A. Schaarschuch (Hrsg.), Soziale Dienstleistungen aus Nutzersicht: Zum Gebrauchswert Sozialer Arbeit (S. 9–25). München: Ernst Reinhardt Verlag.

- Scheirer, M. A. (1978). Program Participants' Positive Perceptions: Psychological Conflict of Interest in Social Program Evaluation. *Evaluation Review*, 2 (1), 53–70.
- Schellberg, K. (2015). Der Social Return on Investment: Strategische Möglichkeiten für den Sozialbereich? In K.P. Sprinkart (Hrsg.), *Nachhaltigkeit messbar machen. Integrierte Bilanzierung für Wirtschaft, Sozialwirtschaft und Verwaltung* (1. Auflage, S. 113–137). Regensburg: Walhalla Verlag.
- Scherenberg, V. (2018). *Gesundheitsökonomische Evaluationen kompakt: für Studium, Prüfung und Beruf (Methodenbuch)* (3. Auflage.). Bremen: APOLLON University Press.
- Schmidt, S. (2016). Nutzen einer wirkungsorientierten Zielsystematik. Bindeglied zwischen Evaluation und Programmsteuerung. In S. Giel, K. Klockgether & S. Mäder (Hrsg.), *Evaluationspraxis: Professionalisierung – Ansätze – Methoden* (2. Auflage, S. 159–175). Münster: Waxmann Verlag.
- Schmitt, J. & Beach, D. (2015). The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, 21 (4), 429–447. doi:10.1177/1356389015607739
- Schober, C. & Then, V. (Hrsg.). (2015). *Praxishandbuch Social Return on Investment: Wirkung sozialer Investitionen messen* (1. Auflage). Stuttgart: Schäffer-Poeschel Verlag.
- Schöffski, O. & Schulenburg, J.-M. (Hrsg.). (2012). *Gesundheitsökonomische Evaluationen* (4., vollständig überarb. Aufl.). Berlin: Springer.
- SECO (Hrsg.). (2017). *Logical Framework – User Manuel*. Staatssekretariat für Wirtschaft SECO Wirtschaftliche Zusammenarbeit und Entwicklung der Schweizerischen Eidgenossenschaft. Zugriff am 28.6.2018. Verfügbar unter: <https://www.seco-cooperation.admin.ch/dam/secocoop/de/dokumente/resultate/monitoring/quali-manual-framework.pdf.download.pdf/Logical%20Framework%20-%20User%20Manual.pdf>
- Sedmak, C., Babic, B., Bauer, R. & Posch, C. (Hrsg.). (2011). *Der Capability-Approach in sozialwissenschaftlichen Kontexten: Überlegungen zur Anschlussfähigkeit eines entwicklungspolitischen Konzepts* (1. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference* (1. Auflage). Belmont: Wadsworth Cengage Learning.
- Shaw, S. & Volz, U. (2017, Januar 1). Was ist Wirkung. Zugriff am 10.5.2018. Verfügbar unter: <http://www.benckiser-stiftung.org/de/blog/what-is-impact>
- Social Reporting Initiative e.V. (2014). *SRS Social Reporting Standard: Leitfaden zur wirkungsorientierten Berichterstattung*. Zugriff am 15.5.2018. Verfügbar unter: http://www.social-reporting-standard.de/fileadmin/redaktion/downloads/SRS_Leitfaden_2014_DE.pdf
- Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung: Eine Grundlage für wirkungsorientiertes Qualitätsmanagement* (1. Auflage). Münster: Waxmann Verlag GmbH.
- The Center for Theory of Change. (o. J.). *How Does Theory of Change Work? Theory of Change Community*. Zugriff am 8.6.2018. Verfügbar unter: <http://www.theoryofchange.org/what-is-theory-of-change/how-does-theory-of-change-work/>
- The SROI Network (Hrsg.). (2012). *A guide to Social Return on Investment*. Zugriff am 22.8.2016.

- Verfügbar unter:
<http://www.socialvalueuk.org/app/uploads/2016/03/The%20Guide%20to%20Social%20Return%20on%20Investment%202015.pdf>
- Vaessen, J. (2016). How complicated does the (Intervention) Model have to be? Zugriff am 7.6.2018. Verfügbar unter: <https://ieg.worldbankgroup.org/blog/how-complicated-does-intervention-model-have-be>
- Veerman, J. W. & van Yperen, T. A. (2007). Degrees of freedom and degrees of certainty: A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning*, 30 (2), 212–221. doi:10.1016/j.evalprogplan.2007.01.011
- Wingenfeld, K., Kleina, T., Franz, S., Engels, D., Mehlan, S. & Engel, H. (2011). Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität in der stationären Altenhilfe. Abschlussbericht. Berlin: Bundesministerium für Gesundheit & Bundesministerium für Familie, Senioren, Frauen und Jugend. Zugriff am 15.5.2018. Verfügbar unter:
<https://www.bmfsfj.de/blob/93206/2dda7f65c418478da3260d2f7996daa2/abschlussbericht-stationaere-altenhilfe-data.pdf>
- Zängl, P. (2013). Impact. In P. Grundwald, G. Horcher & B. Maelicke (Hrsg.), *Lexikon der Sozialwirtschaft* (2. Auflage, S. 471–472). Baden-Baden: Nomos Verlagsgesellschaft.

Bisher erschienene Beiträge:

1. Brendebach, Christine: Die Bedeutung bürgerschaftlichen Engagements angesichts der demographischen Herausforderungen
2. Sommer-Himmel, Roswitha: Wohin bilden wir unsere Kinder? Eltern und Kita unter Druck – wenn Anforderungen und Erwartungshaltungen kollidieren
3. Kranenpohl, Uwe: Die neue Grundordnung der Evangelischen Hochschule Nürnberg
4. König, Joachim: Nachhaltigkeit in der Sozialen Arbeit – Konzeptionelle, praktische und empirische Implikationen aus pädagogischer Sicht
5. Städtler-Mach, Barbara: Grenzen und Verletzlichkeit im Alter
6. Füglein, Kurt: Hochschule ist anders
7. Schellberg, Klaus: Von der Pionierzeit zur Konsolidierung – ein Abriss der Entwicklung des Sozialmanagements
8. Kaltschmidt, Corinna: Habe Fragen, suche Antworten! Die Geschwisterbeziehung in Familien mit Kindern ohne und mit Behinderung
9. Kruse Jürgen: Stationen eines akademischen Lebens als Spiegel gesellschaftlicher Veränderungen
10. Schübler, Marion: Hochschuldidaktik im Kontext der Theaterpädagogik
11. Kranenpohl, Uwe: Keine „Stunde der Exekutive“(?) – Bundestag und Bundesverfassungsgericht in der „Eurokrise“
12. Frisch, Ralf: Gewalt als Krise der Religion – Eine theologische Auseinandersetzung mit der dunklen Seite der Macht
13. Kuch, Michael: Hören und Verstehen – Wodurch das Erkennen Freude macht. Theologische Bemerkungen
14. König, Joachim: Arbeitslosigkeit – Belastungs- und Bewältigungsprozesse als Herausforderung für die Erwachsenenbildung
15. Schübler, Marion: Erfahrungsorientierte Didaktik als Etüde. Methoden als Bedingung für ästhetische Bildungsprozesse

16. Köhler, Anne-Sophie & König, Joachim: Marginalisierte und schwer erreichbare junge Menschen mit komplexen Problemlagen als Zielgruppe der Jugendsozialarbeit
17. Winkler, Kathrin: Migration und Mehrsprachigkeit – Ein kritisches Verhältnis im Bildungskontext Schule
18. König, Joachim: Bildung ganzheitlich denken und gemeinsam verantworten
19. Sommer-Himmel, Roswitha: Akademisierung als Mehrwert in Kindertageseinrichtungen?
20. Sommer-Himmel, Roswitha & Link, Marita: Forschendes Lernen am Beispiel des Lehr- und Lernformates „Praxisforschung“: Eine systematische Begegnung zwischen Theorie und Praxis
21. König, Joachim & Ottmann, Sebastian: Marktforschung in der Sozialwirtschaft – von der Theorie zur Praxis
22. Manzeschke, Arne: Homo imagines faber – Menschenbildlichkeit zwischen Idolatrie und Selbstreflexivität
23. Bayer, Michael: Die Welt aus den sozialen Fugen – Ein soziologischer Blick
24. Wölfel, Anne: Ich kann nicht mehr und jetzt? – Pflegende Angehörige an der Grenze zur Überlastung – Konzeption eines Fragebogens zur Selbsteinschätzung des Belastungsniveaus und der Bewältigungsstrategien, zur Bedarfsermittlung im Rahmen der Beratung pflegender Angehöriger
25. Bauer, Kristina: Professionelle Responsivität der Fachkräfte im Umgang mit konflikthaften Peerinteraktionen in der Krippe
26. Zeus, Andrea: Die Umsetzung des Anspruchs auf inklusive Bildung – ein Thema auch für die Jugendsozialarbeit an Schulen in Nürnberg?! Ein Arbeitspapier
27. Appel, Michael & Strehl, Maximilian: „Ankommen in der Fremde“ – Zentrale Befunde einer biografieanalytischen Studie zu Ankommens- und Integrationsprozessen ehemaliger, minderjähriger Flüchtlinge
28. Frisch: Ist das zu glauben? – Einige schöpfungstheologische Gedanken von womöglich weltbewegender Bedeutung